Aalto University
School of Science
Degree Programme in
Engineering Physics and Mathematics

Lauri Himanen

Hybrid Quantum Mechanical and Molecular Mechanical Modeling in Ion-Water Solutions

Master's Thesis
Espoo, 29th of April, 2015

Supervisor: Professor Adam Foster, Aalto University
Advisor: Professor Adam Foster, Aalto University

Aalto University
School of Science
Degree Programme in
Engineering Physics and Mathematics

ABSTRACT OF
MASTER'S THESIS

Author: Lauri Himanen

Title:
Hybrid Quantum Mechanical and Molecular Mechanical Modeling in Ion-Water Solutions

Date: 29th of April, 2015 **Pages:** 61

Major: Engineering Physics **Code:** Tfy-105

Supervisor: Professor Adam Foster

Advisor: Professor Adam Foster

The methods needed in order to successfully perform hybrid quantum- and molecular mechanical modeling of ion-water solutions are introduced and tested. The relevant theoretical methods are introduced at a general level and are then employed in realistic simulations of infinitely dilute ion-water solutions.

The test simulations are performed with the CP2K simulation software for which a custom object-oriented interface was written. The QM/MM results are compared against equivalent classical simulations and previous experimental and computational results. The quantitative comparison is based on the properties of the first hydration layer, such as position, coordination number, mean residence time and O-Ion-O angular distribution. The possible difference between mechanical and electrical embedding is also examined.

Two biologically important ions, sodium and potassium are used as test systems, but the presented methods are transferable to any other ionic fluids or more complex systems.

Keywords: QM/MM, ion-water solutions, CP2K, radial distribution function, hydration layer, molecular dynamics, DFT, electrical embedding, mechanical embedding, GEEP, adaptive thermostats, adaptive buffered force-mixing

Language: English

Aalto-yliopisto
Perustieteiden korkeakoulu
Teknillisen fysiikan ja matematiikan koulutusohjelma

DIPLOMITYÖN
TIIVISTELMÄ

Tekijä:	Lauri Himanen		
Työn nimi:	Yhdistetty kvanttimekaaninen ja molekyylimekaaninen mallintaminen vesi-ioni-liuoksissa		
Päiväys:	29.05.2015	**Sivumäärä:**	61
Pääaine:	Teknillinen fysiikka	**Koodi:**	Tfy-105
Valvoja:	Professori Adam Foster		
Ohjaaja:	Professori Adam Foster		

Työssä tarkastellaan yhdistetyn kvanttimekaanisen ja molekyylimekaanisen mallintamisen (QM/MM-mallinnus) käyttöä vesi-ioni-liuoksien laskennallisessa simuloinnissa. Simulaatioiden toteuttamiseen liittyvä taustateoria käydään päällisin puolin läpi ja tätä teoriaa sovelletaan käytännön simulaatioissa.

Testisimulaatiot toteutetaan CP2K-ohjelmistolla jolle luotiin työn aikana myös oliopohjaisen ohjelmoinnin mukainen käyttöliittymä. QM/MM-mallinnuksen avulla saatuja tuloksia verrataan vastaavaan klassiseen simulaatioon, sekä aikaisempiin kokeellisiin ja laskennallisiin tuloksiin. Kvantitatiivinen vertailu tehdään vertaamalla ensimmäisen solvaattiverhon ominaisuuksia, kuten solvaattiverhon sijaintia, sen sisältämien vesimolekyylien määrää, vesimolekyylien pysyvyyttä verhossa sekä verhossa esiintyviä kahden happiatomin ja ionin määrämiä kulmia. Myös mekaanisen ja sähköisen sulauttamisen eroja QM/MM-mallissa tutkitaan.

Kahta biologisesti tärkeää ionia, natriumia ja kaliumia käytetään esimerkkisysteemeinä. Esitellyt menetelmät ovat kuitenkin siirrettävissä myös muiden ionien mallintamiseen tai myös monimutkaisempiin systeemeihin.

Asiasanat:	QM/MM, vesi-ioni-liuokset, CP2K, radiaalinen jakaumafunktio, solvaattiverho, molekyylidynamiikka, tiheysfunktionaaliteoria, sähköinen sulauttaminen, mekaaninen sulauttaminen, GEEP, mukautuvat termostaatit, adbf
Kieli:	Englanti

Acknowledgements

Espoo, 29th of April, 2015

Lauri Himanen

Abbreviations and Acronyms

QM	Quantum Mechanical
MM	Molecular Mechanical
RDF	Radial Distribution Function
adbf	Adaptive Buffered Force Mixing
DFT	Density Functional Theory
ME	Mechanical Embedding
EE	Electrical Embedding
GEEP	Gaussian Expansion of the Electrostatic Potential
HF	Hartree–Fock

Contents

Chapter 1

Introduction

The modeling of atomistic systems can be roughly divided into two main categories: classical molecular mechanical modeling and quantum mechanical modeling. In the classical models atoms or groups of atoms are the fundamental particles of interest and the electron behaviour is only implicitly described. Classical models can be used to describe even hundreds of millions of atoms [1], and they have given much insight into the structure of matter. The next step in accuracy is to explicitly model the electron behaviour which is key to all chemical bonding. Unlike the nucleus, the electrons in atoms require a quantum mechanical description due to them having much less mass and being more weakly bound than nucleons. The explicit quantum mechanical modeling of the electrons presents a huge leap in the accuracy of the models, but also in the computational complexity.

Sometimes the system can be logically divided into two separate regions which can be modeled with different levels of theory: classical modeling for one, quantum modeling for the other. In this way one can still capture the essential physics of the system without paying high computational price of doing a full quantum mechanical calculation. This is called QM/MM modeling.

This mixing of classical and quantum mechanical models in computational nanoscience has a long tradition which dates back to the 70'ies. Although the computational power has developed according to Moore's law, so has the ambition of the research done in this field. This is why the hybrid methods will stay relevant, although the meaning of higher and lower level of theory might change in this context. The early developers of these methods, Martin Karplus, Michael Levitt and Arieh Warshel, were awarded the 2013 Nobel Prize in Chemistry [2].

Ion-water solutions present an interesting environment for the hybrid methods, as one might want to capture the effect of the ion in the system

quantum mechanically. This can be achieved by modeling the ion and it's immediate surroundings with a higher level theory and rest of the system with a lower level classical theory. Ion-water solutions are found in many applications, and are almost always present in biological systems, being responsible for many crucial biophysical functions such as nerve and muscle excitation.

1.1 Motivation and Goals

The goal of this work is to get familiar with the theoretical framework behind QM/MM modeling and to employ these theoretical tools in computational simulations of ion-water systems. The success of the QM/MM model is based on it's ability to produce realistic results in feasible computational time. A direct comparison between QM/MM embedding schemes is performed to investigate whether the more complex electrical embedding is needed in the investigated systems.

The computational study is limited to two biologically relevant alkali ions, Na^+ and K^+, but the presented methods are transferable to other similar systems. The presented simulation methods can thus be used for various different ionic species, or for more complex systems.

1.2 Overview of Contents

In chapter 2 the background of ion-water solutions and the previous experimental and computational results are introduced. Chapter 3 is dedicated to presenting the theory behind each of the three models: MM, QM and QM/MM. In chapter 4 the object-oriented scripting environment PYCP2K that was developed during this work is briefly introduced. The simulation and analysis methods are presented in chapter 5. The results are presented in chapter 6 and the discussion and conclusions are left for chapter 7.

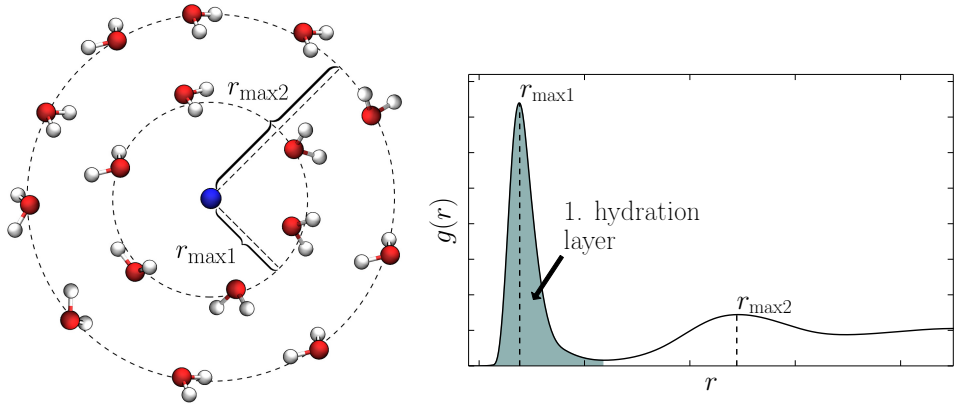
Chapter 2

Ion-Water Solutions

Ions are ubiquitous in nature and have a fundamental role in numerous fields. They have a key role in biochemical and physiological phenomena (membrane potentials, nerve action potentials and muscle contraction, stabilization and modulation of biomolecular structure, intracellular signal transduction, enzyme and nucleic acid catalysis), but also in numerous ecological (geology, atmospheric chemistry) and technical (electrolysis, batteries, materials engineering) processes [3].

Because of the magnitude and range of the electrostatic interactions between ions, they tend to form ionic salts. In water however, the highly polar water molecules are able to dissolve these salts, creating isolated ions surrounded by structured layers of water. These hydration layers can be experimentally or computationally analyzed from e.g. the radial distribution function between ions and oxygens in water.

The detailed structure of the hydration layers can be critical especially in biophysical processes. Table 2.1 shows the approximate concentrations of important electrolytes in the extracellular- and intracellular fluid [4]. Ion channels and pumps in the cell membrane control this delicate balance by selectively allowing only one or few different ions or molecules to pass. For example, the potassium transport channel is 3 by 3 ångströms in size, and is not charged so the ions are not pulled into the channels by electrostatic forces [4]. The hydrated form of the potassium ion is considered smaller than the hydrated form of sodium because the sodium ion attracts attracts more water molecules than does potassium [4]. Therefore, the smaller hydrated potassium ion can pass through this small channel, whereas the larger hydrated sodium ions are rejected, thus providing selective permeability for a specific ion.



(a) (b)

Figure 2.1: a) The formation of hydration layers around ions in water. The polar water aligns in spherical layers around an ion. b) Typical radial distribution function between an ion and oxygens in water. The first hydration layer is shaded and it's position is determined by the maximum value r_{max1}. The first layer extends to the minimum between the first and second layers.

Table 2.1: The approximate concentrations of important electrolytes in extracellular fluid and intracellular fluid [4]. Note the concentration differences between e.g. sodium and potassium.

Electrolyte	Extracellular fluid	Intracellular fluid
Na^+	142 mEq/l	10 mEq/l
K^+	4 mEq/l	140 mEq/l
Ca^{++}	2.4 mEq/l	0.0001 mEq/l
Mg^{++}	1.2 mEq/l	58 mEq/l
Cl^-	103 mEq/l	4 mEq/l
HCO_3^-	28 mEq/l	10 mEq/l
SO_4^-	1 mEq/l	2 mEq/l
Phosphates	4 mEq/l	75 mEq/l

2.1 Model Ions Na^+ and K^+

As model ions we chose Na^+ and K^+ which are among the most important metal ions in biology as well as industry. Their importance in biology is reflected by their high abundance in living systems and in nature in general.

Sodium and potassium ions are key constituents of many minerals, and their general high solubility is responsible for the composition of marine water [5]. Abundance, high solubility and reaction inertness are desirable properties in industrial processes where they are frequently used as counterions for anions with desirable properties [5]. Although we use these two ions in our simulations the used methods can very well be extended to other ions as well.

These ions also differently affect the local water-water interactions around them, although they both belong to the same first group in the periodic table and have the same charge. K^+ is seen as "chaotropic" [6] as it can disrupt the hydrogen bonding network between water molecules, thus breaking the water's local structure. Na^+ on the other hand is on the borderline between chaotropic and cosmotropic [6], cosmotropic being the opposite of chaotropic.

2.1.1 Previous Experimental Results

The structural properties of alkali, alkaline earth and halide ions in water have been widely studied using experimental techniques, such as x-ray and neutron diffraction. In theory the diffraction techniques can determine accurate radial distribution functions between the ions and water oxygen atoms. From the radial distribution function one can further identify the position of hydration shells and also the average number of water molecules in a certain hydration shell. Diffraction measurements on liquids and the subsequent data analysis are however particularly challenging and this has presumably caused the relatively large variation in the experimental results as seen in table 2.2.

2.1.2 Previous Simulation Results

Due to their presence in many processes that are hard to realize experimentally, ion-water systems have been subject to a rich variety of computational simulations. There are multiple classical force field parametrizations available for different ions, fully quantum simulations have been performed and a variety of different hybrid schemes have been employed. Some of the computational results are gathered in table 2.3 for reference. The table reveals that there is no prevailing consensus among the results, and that the topic is subject to active research.

Table 2.2: First hydration shell properties of Na^+ and K^+ ions from experiments. r_{max1} refers to the first maximum in the ion–oxygen radial distribution function. N_1 refers to the coordination number of the first hydration shell.

Ion	Method	r_{max1}(Å)	N_1	year	ref
Na	X-ray diffraction	2.42	4.7	1980	[7]
	X-ray diffraction	2.40	4.9	1989	[8]
	Neutron diffraction	2.3	4.9	2006	[9]
	Neutron diffraction	2.34	5.3	2007	[10]
	X-ray diffraction	2.43	6	2011	[11]
K	X-ray diffraction	2.8	6	1980	[7]
	Neutron diffraction	2.9–3.4	5.3	2006	[9]
	Neutron diffraction	2.65	6.0	2007	[10]
	X-ray diffraction	2.81	7	2011	[11]

Table 2.3: First hydration shell properties of Na^+ and K^+ ions from computer simulations. r_{max1} refers to the first maximum in the ion–oxygen radial distribution function. N_1 refers to the coordination number of the first hydration shell. τ_0 and $\tau_{0.5}$ refer to the mean residence times with time constants of 0 and 0.5 ps respectively.

Ion	Method	Water model	r_{max1}(Å)	N_1	τ_0	$\tau_{0.5}$	year	ref
Na	MM	SPC/E	2.38	-	-	-	2008	[12]
	MM	TIP4P$_{EW}$	2.35	-	-	-	2008	[12]
	MM	TIP4P	2.45	6.6	-	-	1994	[13]
	MM	SPC/E	2.45	5.8	-	-	1998	[14]
	QM/MM (HF)	BJH-CF2	2.33	5.6	-	-	1998	[15]
	QM/MM (HF)	BJH-CF2	-	5.4	0.4	2.4	2004	[16]
	QM/MM (HF)	BJH-CF2	2.39	6.0	0.94	7.4	2010	[17]
	QM/MM (DFT)	SWM4-NDP	2.3–2.5	5.4	-	-	2012	[18]
	QM (DFT)	-	2.35	4.6	-	-	2001	[19]
	QM (DFT)	-	2.49	5.2	-	-	2000	[20]
K	MM	SPC/E	2.74	-	-	-	2008	[12]
	MM	TIP4P$_{EW}$	2.72	-	-	-	2008	[12]
	MM	TIP4P	2.70	8.0	-	-	1994	[13]
	MM	SPC/E	2.80	7.2	-	-	1998	[14]
	QM/MM (HF)	BJH-CF2	2.81	8.3	-	-	1998	[15]
	QM/MM (HF)	BJH-CF2	-	8.0	0.3	2.0	2004	[16]
	QM/MM (HF)	BJH-CF2	2.85	8.8	0.47	3.4	2010	[17]
	QM/MM (DFT)	SWM4-NDP	2.7–2.8	6.8	-	-	2012	[18]
	QM/MM (DFT)	BJH-CF2	2.78	6.3	0.42	1.8	2011	[21]
	QM (DFT)	-	2.8	5.9	-	-	2004	[22]

Chapter 3

Theory

In this chapter we review the basic theoretical framework of how the interactions and time-development in an atomistic simulation can be modeled and how these together with the correct approximations and computational techniques can lead to realistic simulations. We start from the simplest classical model, scratch the surface of the quantum models heavily focusing on Born-Oppenheimer Molecular dynamics and density functional theory and finally see how the two models can be combined in the QM/MM models.

3.1 Classical Molecular Mechanics

In classical molecular Dynamics atoms are the smallest particles that are explicitly modeled. Energetic interactions between the atoms are represented by potentials with predefined analytical forms. These potentials are derived from both experimental work and high-level quantum mechanical calculations.

The electrons, which determine all chemical behaviour, are thus only implicitly modeled as classical potential fields. This approximation allows the study of large biological systems or material assemblies with millions of atoms [1]. The typical parameters used to capture the interactions can consist of e.g. bond lengths, bond angles, partial charge values and van der Waals parameters. Different implementations of molecular mechanics use different expressions for the potential functions, but in general the potentials could be broken to the following parts:

$$V_{\text{total}} = V_{\text{covalent}} + V_{\text{noncovalent}} \tag{3.1}$$

$$V_{\text{covalent}} = V_{\text{bond}} + V_{\text{angle}} + V_{\text{dihedral}} + V_{\text{bond-order}} \tag{3.2}$$

$$V_{\text{noncovalent}} = V_{\text{electrostatic}} + V_{\text{van Der Vaals}} \tag{3.3}$$

The potential functions combine to give the total energy of the system, and the force F_i on atom i can be directly calculated from the potential V as

$$\mathbf{F}_i\left(\{\mathbf{r}_M\}\right) = -\nabla_i V\left(\{\mathbf{r}_M\}\right) \tag{3.4}$$

where $\{\mathbf{r}_M\}$ is the set of M atomic coordinates. These forces can be then used to drive the dynamics of the system.

Despite their success, the need to device a fixed predefined potential implies certain drawbacks. Crafting the potentials can be very tedious and due to their fixed nature their explanatory power is often restricted to certain predefined scenarios. This is why molecular mechanics is typically unable to describe processes that involve complex bond-breaking and bond-forming, charge transfer, and electronic excitation and thus require quantum mechanics for a proper treatment [23].

3.2 Quantum Mechanical Methods

The foundation behind quantum mechanical electronic structure calculations is the non-relativistic time-dependent Shrödinger equation

$$i\hbar\frac{\partial}{\partial t}\Phi\left(\{\mathbf{r}_i\},\{\mathbf{R}_I\};t\right) = \hat{H}\Phi\left(\{\mathbf{r}_i\},\{\mathbf{R}_I\};t\right) \tag{3.5}$$

with the Hamiltonian operator

$$\begin{aligned}\hat{H} = &-\sum_I \frac{1}{2M_I}\nabla_I^2 - \sum_i \frac{1}{2m_\mathrm{e}}\nabla_i^2 \\ &+ \frac{1}{4\pi\epsilon_0}\sum_{i<j}\frac{e^2}{|\mathbf{r}_i-\mathbf{r}_j|} - \frac{1}{4\pi\epsilon_0}\sum_{I,i}\frac{e^2 Z_I}{|\mathbf{R}_I-\mathbf{r}_i|} + \frac{1}{4\pi\epsilon_0}\sum_{I<J}\frac{e^2 Z_I Z_J}{|\mathbf{R}_I-\mathbf{R}_J|}\end{aligned} \tag{3.6}$$

The total wave function Φ depends on the electronic $\{\mathbf{r}_i\}$ and the nuclear $\{\mathbf{R}_I\}$ degrees of freedom, including spin. M_I and Z_I are the mass and atomic number of the Ith nucleus, the electron mass and elementary charge are denoted by m_e and e and ϵ_0 is the vacuum permittivity.

There are multiple ways of approximately solving this equation and determining the energies and time-development of the quantum system, including Born-Oppenheimer, Ehrenfest and Car-Parrinello molecular dynamics [24]. In this work we will take the approach of Born-Oppenheimer Molecular Dynamics which is discussed next.

3.2.1 Born-Oppenheimer Molecular Dynamics

The problem of solving a time-dependent many–body quantum system with nuclear wavefunctions can luckily be simplified with certain approximations. Even the lightest of all nuclei, the proton, weighs roughly 10^3 times more than an electron, and for a typical nucleus such as carbon the mass ratio well exceeds 10^4. Thus, the nuclei move much slower than the electrons. The practical consequence is that we can – at least to a good approximation – take the extreme point of view and consider the electrons as moving in the field of fixed nuclei [25]. This is the so called "clamped nuclei approximation". With this approximation the kinetic energy of the nuclei can be disregarded in equation (3.6) and the positions of the nuclei enter only as fixed parameters. The nucleus–nucleus interaction also reduces to a mere constant. The Hamiltonian in equation (3.6) then becomes

$$\hat{H}_{\text{elec}} = -\sum_i \frac{1}{2m_{\text{e}}}\nabla_i^2 + \frac{1}{4\pi\epsilon_0}\sum_{i<j}\frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} - \frac{1}{4\pi\epsilon_0}\sum_{I,i}\frac{e^2 Z_I}{|\mathbf{R}_I - \mathbf{r}_i|} + \underbrace{\frac{1}{4\pi\epsilon_0}\sum_{I<J}\frac{e^2 Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|}}_{=\text{constant}} \tag{3.7}$$

This Hamiltonian now describes the behaviour of an electronic wave function $\psi_k\left(\{\mathbf{r}_i\};\{\mathbf{R}_I\}\right)$, which depends on the nucleic positions only parametrically. The second approximation is to restrict this wave function to the ground state. This approximation should be valid as long as the difference between the ground state energy and the first excitation energy is large everywhere compared to the thermal energy $k_{\text{B}}T$ [24]. The adiabatic theorem [26] then ensures that the electronic wave function will stay in the ground state, as the electron-nuclei potential slowly (from the electronic point of view) changes. This means that we can concentrate on the time-independent electronic Shrödinger equation:

$$\hat{H}_{\text{elec}}\psi_k\left(\{\mathbf{r}_i\};\{\mathbf{R}_I\}\right) = E_k\psi_k\left(\{\mathbf{r}_i\};\{\mathbf{R}_I\}\right)) \tag{3.8}$$

Now that the system behaviour from the electron's perspective has been determined, we need to consider the system behaviour from the perspective of the nuclei. Being much more massive and highly localized, the nuclei can be modeled classically and they move in the potential surface $E_k\left(\{\mathbf{R}_I\}\right)$ obtained by solving the equation (3.8) with the given nuclear configuration

$$\mathbf{F}_I = -\nabla_I E_k\left(\{\mathbf{R}_I\}\right) \tag{3.9}$$

Another, less computationally expensive way to calculate the intramolecular forces is to employ the Hellmann–Feynman theorem which states that the force F_{x_i} for the ith nucleus in the direction of the coordinate x_i is given by

$$F_{x_i} = -\int \psi^* \frac{\partial \hat{H}}{\partial x_i} \psi \mathrm{d}\mathbf{r} = -\int \psi^* \psi \frac{\partial V}{\partial x_i} \mathrm{d}\mathbf{r} \tag{3.10}$$

where ψ is an eigenstate of the hamiltonian $\hat{H} = \hat{T} + V$, where the kinetic energy operator $\hat{T}$ of the system has been separated as it doesn't affect the forces.

The core of the problem has thus been reduced to equation (3.8), which is still a many-body problem. Next we will look at one way of efficiently solving this equation by using density functional theory.

3.2.2 Density Functional Theory

Density functional theory (DFT) as we know it today was born in 1964 when a landmark paper by Hohenberg and Kohn appeared in the Physical Review [25]. The two Hohenberg–Kohn theorems proven in this report represent the major theoretical pillars on which all modern day density functional theories are based on. The first Hohenberg–Kohn theorem states that the electron density $\rho(\mathbf{r})$ for N electrons

$$\rho(\mathbf{r}) = N \int \dots \int |\psi\left(\{\mathbf{r}_\mathrm{i}\}\right)|^2 \mathrm{d}s_1 \mathrm{d}\mathbf{r}_2 \dots \mathrm{d}\mathbf{r}_\mathrm{N} \tag{3.11}$$

– a quantity that depends only on the three spatial variables – uniquely determines the ground state wave function and thus all the properties of the ground state system [25]. The second Hohenberg–Kohn theorem states that for any trial density $\rho_\mathrm{t}(\mathbf{r})$ for which $\int \rho_t(\mathbf{r}) \mathrm{d}\mathbf{r} = N$ and which is associated with some nuclear coordinates and charges, the following variational principle holds

$$E[\rho_\mathrm{t}(\mathbf{r})] \geq E_0 \tag{3.12}$$

where E_0 is the true ground state energy and $E[\rho_\mathrm{t}(\mathbf{r})]$ is the energy functional related to the trial density [25].

Based on these two theorems it is justified to do variational calculations with electron density in order to search for the ground state of a system. However these theorems do not provide any guidance at how the functional $E[\rho_\mathrm{t}(\mathbf{r})]$ that delivers the ground state energy should be constructed. This problem is discussed next, as the Kohn-Sham approach is introduced.

3.2.2.1 Energy functional in Kohn-Sham approach

In 1965 [25] Kohn and Sham proposed a way to construct the energy functional so that it produces physically reasonable solutions and can be easily translated to a computational task. Nowadays basically all modern DFT codes are based on this approach. The first step is to extract the known electron-electron and electron-nuclei Coulomb interactions. This separation can be written as

$$E\left[\rho\right] = E_{\mathrm{C}}\left[\rho\right] + E'\left[\rho\right] \tag{3.13}$$

where the Coulomb term is defined as

$$E_{\mathrm{C}}\left[\rho\right] = E_{\mathrm{C,\,Me}} + E_{\mathrm{C,\,ee}} \tag{3.14}$$

$$= \sum_I \int \frac{\rho\left(\mathbf{r}\right) Z_I}{4\pi\epsilon\left|\mathbf{r}-\mathbf{R}_I\right|}\mathrm{d}\mathbf{r} + \int\int \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{4\pi\epsilon\left|\mathbf{r}_1-\mathbf{r}_2\right|}\mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2 \tag{3.15}$$

The electron-electron interaction actually contains "self-interaction" as the electrons are allowed to interact with themselves as more easily seen from the one electron case of hydrogen where $E_{\mathrm{C,\,ee}} \neq 0$. Correcting this self-interaction problem is left for $E'\left[\rho\right]$.

The insight of Kohn and Sham was to also separate a part of the kinetic energy of the electrons with an approximation inspired by the Hartree-Fock model. The wave function is expanded as an antisymmetric product of spin orbitals ϕ

$$\Psi\left(\{\mathbf{r}_i\}\right) = \frac{1}{\sqrt{N}}\begin{vmatrix} \phi_1(\mathbf{r}_1) & \phi_2(\mathbf{r}_1) & \cdots & \phi_N(\mathbf{r}_1) \\ \phi_1(\mathbf{r}_2) & \phi_2(\mathbf{r}_2) & \cdots & \phi_N(\mathbf{r}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(\mathbf{r}_N) & \phi_2(\mathbf{r}_N) & \cdots & \phi_N(\mathbf{r}_N) \end{vmatrix} \tag{3.16}$$

This wave function is the eigenfunction of an Hamiltonian

$$\hat{H}_{\mathrm{S}} = -\frac{1}{2m_{\mathrm{e}}}\sum_i^N \nabla_i^2 + \sum_i^N V_{\mathrm{eff}}(\mathbf{r}_i) \tag{3.17}$$

which corresponds to a system of non-interacting electrons in an effective local potential $V_{\mathrm{eff}}(\mathbf{r})$. Although this system does not correspond to the real one, it's electron density can be matched to the real system by determining the correct effective potential. Then the kinetic energy of this non-interacting system can be calculated as

$$E_{\mathrm{K}} = -\frac{1}{2m_{\mathrm{e}}}\sum_i^N \left\langle \phi_i \left| \nabla_i^2 \right| \phi_i \right\rangle \tag{3.18}$$

Of course, the non-interacting kinetic energy is not equal to the true kinetic energy of the interacting system, even if the systems share the same density. Nevertheless, in this way a major part of the kinetic energy can be calculated. The energy functional can now be separated as follows

$$E\left[\rho\right] = E_{\mathrm{C}}\left[\rho\right] + E_{\mathrm{K}}\left[\rho\right] + E_{\mathrm{XC}}\left[\rho\right] \tag{3.19}$$

where E_{XC} is the so–called exchange-correlation functional that contains everything that is yet unknown. In spite of its name, it also contains the remaining kinetic energy and the self-interaction correction in addition to the exchange and correlation energies.

By applying the variational principle and minimizing the energy with respect to the spin orbitals with the constraint of $\langle\phi_i|\phi_j\rangle = \delta_{ij}$, one can derive the Kohn-Sham equations for the orbitals ϕ_i [25]

$$\left[-\frac{1}{2}\nabla^2 + \left(e^2\int\frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\mathrm{d}\mathbf{r}' + \frac{\delta E_{\mathrm{XC}}[\rho]}{\delta\rho(\mathbf{r})} - \sum_I^M \frac{eZ_I}{|\mathbf{r}-\mathbf{R}_I|}\right)\right]\phi_i \tag{3.20}$$

$$= \left(-\frac{1}{2}\nabla^2 + V_{\mathrm{eff}}(\mathbf{r})\right)\phi_i = \epsilon_i\phi_i \tag{3.21}$$

where V_{eff} already depends on the orbitals and therefore these equations need to be solved self-consistently through iterations. If the exact form of E_{XC} was known, the Kohn–Sham strategy would lead to the exact ground-state energy. The approximation only enters when we have to decide on an explicit form of the unknown functional for the exchange-correlation E_{XC}. The central goal of modern density functional theory is therefore to find better and better approximations to this quantity.

To conclude, in density functional theory we avoid explicitly solving the highly complicated Schrödinger equation (3.8) by using the electron density as a central quantity of interest. The exchange-correlation functionals used in the real-world DFT applications are always approximations, and DFT typically has problems describing e.g. excited states, highly degenerate systems and weak interactions [25]. Nevertheless DFT has enjoyed wide success and continues to evolve.

3.3 Hybrid QM/MM Methods

The starting point in hybrid QM/MM methods is the realization that various regions in the system often play very different roles. Often it is necessary to employ the highest level of theory only in certain most reactive parts

of the system, while other regions may be modeled with simpler, and less computationally expensive techniques. The criteria for choosing QM/MM modeling could be summarized as follows:

- **With full quantum modeling you cannot complete the simulation you're interested in the time-span that is given to you**. This criterion is more practical in nature, as you will have a certain amount of time and computational resources that limit your system.
- **You can identify a smaller part of your system, which must be modeled with quantum methods in order to describe it properly.** Typically this is a part where chemical reaction are supposed to happen, or where your classical model otherwise fails to satisfactorily explain the true system behaviour. You must also ensure that this part is now small enough considering the resource limitations mentioned in the previous criterion. Choosing this part involves knowledge about the physical behaviour of you system.
- **You know that the environment will interact in a meaningful way with the primary QM system, but which can be itself modeled classically.** This part will become the secondary MM subsystem. This criterion is also more physical, as you will have to know how the surroundings interact with the key part of your simulation. The secondary subsystem interacts directly with the QM subsystem via covalent, electrostatic or other non-bonded interactions but also in a larger scale through structure.

The incorporation of quantum mechanics into molecular mechanics can be accomplished in various ways, and one of them is the so-called combined quantum mechanical and molecular mechanical (QM/MM) methodology [23]. A QM/MM method treats a localized region in the system with QM methods and the rest of the system with MM methods. One may also allow atoms to switch description during dynamics, which will be discussed in section 3.3.3

Broadly speaking, the hybrid models can be classified according to whether they combine the classical and quantum mechanical models on the level of energies or forces. Those that build a combined total energy function can do so additively (by combining the QM energy, the MM energy and an interaction energy), subtractively (by combining the MM energy of the entire system with the difference between the QM and MM energies of the QM region) or by mixing local energies computed by QM and MM [27]. Hybrid approaches that are based on forces, on the other hand, combine QM forces for atoms in the QM region with MM forces in the MM region, possibly with

an interpolation between the two in a transition region. The differences between these two coupling schemes become most apparent when the QM and MM partitioning is allowed to change, and the comparison is left for section 3.3.3.

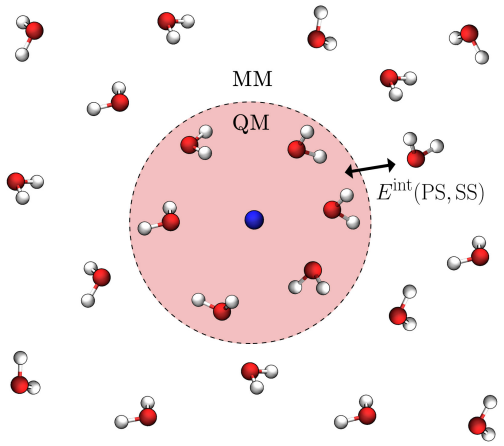


Figure 3.1: System schematic for a simple QM/MM hybrid system. The full atomic structure is divided into QM and MM subsystems and the interaction between them is modeled with a certain level of accuracy.

The inclusion of the interaction between the quantum and classical regions allows a realistic description of the system, and is the core of a QM/MM method. The coupling, in general, must be capable of treating both covalent and non-covalent interactions. Some of the interactions can be reproduced at the MM level, such as van Der Waals interactions, other interactions may need more sophisticated quantum descriptions. Especially tricky are the covalent bondings between the regions and the electrostatic interaction, which are next further elaborated.

3.3.1 Non-covalent Interactions

The non-covalent interactions between QM and MM typically include a Lennard-Jones type of interactions to model the van Der Waals forces and the electron repulsion and a long range electrostatic Coulomb interaction. The electrostatic interaction is typically much more complex of these two, and is elaborated in the following subsection.

3.3.1.1 Electrostatic interaction

The treatment of electrostatic interactions between subsystems in the QM/MM methodology can be divided into three groups [28] with different levels of complexity and physical accuracy.

The first and simplest method is mechanical embedding (ME), in which the QM electron density is not polarized by the charges is the classical region, and the electrostatic interaction between atoms in the subsystems is calculated at the MM level [23] as interaction between point charges.

In the second method, called electrical embedding (EE), the QM computation is carried out in the presence of the partial charges of the classical region by including an additional coulomb interaction term to the Coulomb energy functional (3.15)

$$E_{\mathrm{C,\ MM\text{–}QM}}[\rho] = \int d\vec{r}\rho_{\mathrm{QM}}(\vec{r})V_{\mathrm{MM}}(\vec{r}) \tag{3.22}$$

where $V_{\mathrm{MM}}(\vec{r})$ is the Coulomb potential produced by the classical partial charges in the secondary MM system, and ρ_{QM} is the electron density in the primary QM subsystem. Essentially this means that the quantum electron density gets polarized by the MM charges. The electric embedding may cause overpolarization of the QM electron density known as electron "spill-out". This effect results from the lack of electronic repulsion between the QM electrons and the MM partial charges, as the explicit electrons are missing from the MM model. This effect is particularly pronounced when using a plane-wave basis for the QM electrons.

The third and most refined methods known as self-consistent embedding schemes, refine the electrical embedding scheme by also taking into account the polarization of the MM region in the presence the electric field generated by the QM region [23]. However, this scheme requires a polarizable MM potential, which has the flexibility to respond to perturbation by an external electric field. Such flexibility is not available in most of today's popular MM potentials [23].

3.3.1.2 Electrical Embedding in CP2K

In the CP2K [29] simulation software package, which is used for the simulations in this work, a highly efficient and accurate method called GEEP (Gaussian Expansion of the Electrostatic Potential) is employed for calculating the electrostatic embedding term of equation (3.22). Advanced methods are needed because in QM/MM simulations the calculation of electrical embedding can require between 20-100% of the time required by the QM

calculation, even when using sophisticated hierarchical multipole methods or electrostatic cutoffs [30].

The idea is to first modify the Coulomb interaction term by representing it with a non-diverging but still long ranged potential

$$\frac{1}{r} \approx \frac{Erf\left(\frac{r}{r_c}\right)}{r} \tag{3.23}$$

This modification is justified, because it reduces the unphysical overpolarization in the QM region generated due to the diverging MM point charges. This now very smooth function can be then expanded with a series of Gaussian functions and a very smooth, long-ranged residual term [30]

$$\frac{Erf\left(\frac{r}{r_c}\right)}{r} = \sum_{N_G} A_G e^{-\left(\frac{r}{G_g}\right)^2} + R_{low}(r) \tag{3.24}$$

The A_g are the amplitudes of the Gaussian functions, and G_g are their width.

These different terms can then be evaluated on an efficient multigrid system. Each term in this expression is assigned to a computational grid with different densities. The more flat the term in the expansion, the less dense grid is needed in order to capture it's behaviour. The gaussian terms also have a finite extent, thus they only contribute within a certain cutoff radius. The residual term extends basically all over the simulation cell, but can be evaluated on the coarsest grid due to it being very smooth, provided that the A_g and G_g parameters are properly chosen.

In this way only a fraction of the electrostatic interaction need to be evaluated on the finest grid corresponding to the electron density grid and most terms can be evaluated on a much coarser grid. This results in around a factor of 100 boost in efficiency[30], with essentially no loss in accuracy. This method can also be extended to periodic systems with Ewald summation [31], as the residual term can be efficiently calculated in the reciprocal space with a very low cutoff.

3.3.2 Covalent Bonding

The QM/MM boundary may go through a covalent bond. In such cases, special care is required to treat the boundary. In our ion-water system there is no need to cut any covalent bonds, but in most cases you will not be so lucky. Any even slightly more complex environment will typically have this "bond-cutting", and for this reason we briefly discuss how to deal with it.

Typically it is not sufficient to just model the bonds at the boundary with a MM description. The covalent bonds have a big impact on the configuration of the primary quantum system because electrons have in these cases found a very energetically favourable configuration. If we strip away this covalent environment, the new minimum energy configuration for the electron density might look very different. These changes in the electron density will directly alter the energies in the QM subsystem, but also the QM/MM electrostatic interaction. The secondary system does not suffer from this, because electrons are not explicitly modeled.

There are many ways to deal with bond-cutting, such as local self-consistent field [32], generalized hybrid orbital [33], pseudobonds [34] and link atoms [28]. Here only the simplest and also widely used link-atom approach is discussed in more detail.

3.3.2.1 Link Atoms

In the link atom approach an additional link atom is used to saturate the dangling bond at the QM atom. This link atom is usually taken to be a hydrogen atom [28], and it is augmented to the primary system. The position of the link atom should depend on the coordinates of the covalently bonded QM and MM boundary atoms. Such a constraint removes the extra degrees of freedom due to the link atom. Usually the link atom is put on the line that connects these atoms. The total energy E^{tot} of a QM/MM system with hydrogen link atoms denoted by HL can be formally defined as [28]:

$$E^{\text{tot}} = E^{\text{tot}}_{\text{MM}}(\text{MM}) + E^{\text{tot}}_{\text{QM}}(\text{QM, HL}) + E^{\text{int}}(\text{QM, MM}) + E^{\text{link}} \tag{3.25}$$

$$E^{\text{link}} = -E^{\text{int}}_{\text{MM}}(\text{QM, HL}) - E^{\text{tot}}_{\text{MM}}(\text{HL}) \tag{3.26}$$

where the participating systems are in parenthesis, subscripts indicate the energy calculation method and the superscripts indicate whether the energy is the total energy or the interaction energy of the participating systems.

The E_{link} is the so called link atom correction energy [28]. The first correction term $E^{\text{int}}_{\text{MM}}(\text{QM, HL})$ comes from the fact that the purpose of the hydrogen links is to to alter the electron distribution of the primary system, not to simulate the binding between QM and MM. This is why we have to remove the interaction energy between link atoms and the primary system. The second correction term $E^{\text{tot}}_{\text{MM}}(\text{HL})$ removes the interaction between link atoms themselves, which is counted twice.

It should be kept in mind that the equation (3.25) is only approximately valid, since it assumes that the link interactions in the QM region can be canceled with the MM description. Any big differences in the QM and MM

energies may cause serious errors in energy optimization and force calculations.

The link atom method is straightforward and is widely used. However, it introduces the artificial link atoms that are not present in the original molecular system, and this makes the definition of the QM/MM energy more complicated.

3.3.3 Adaptive QM/MM Methods and Interfacial Errors

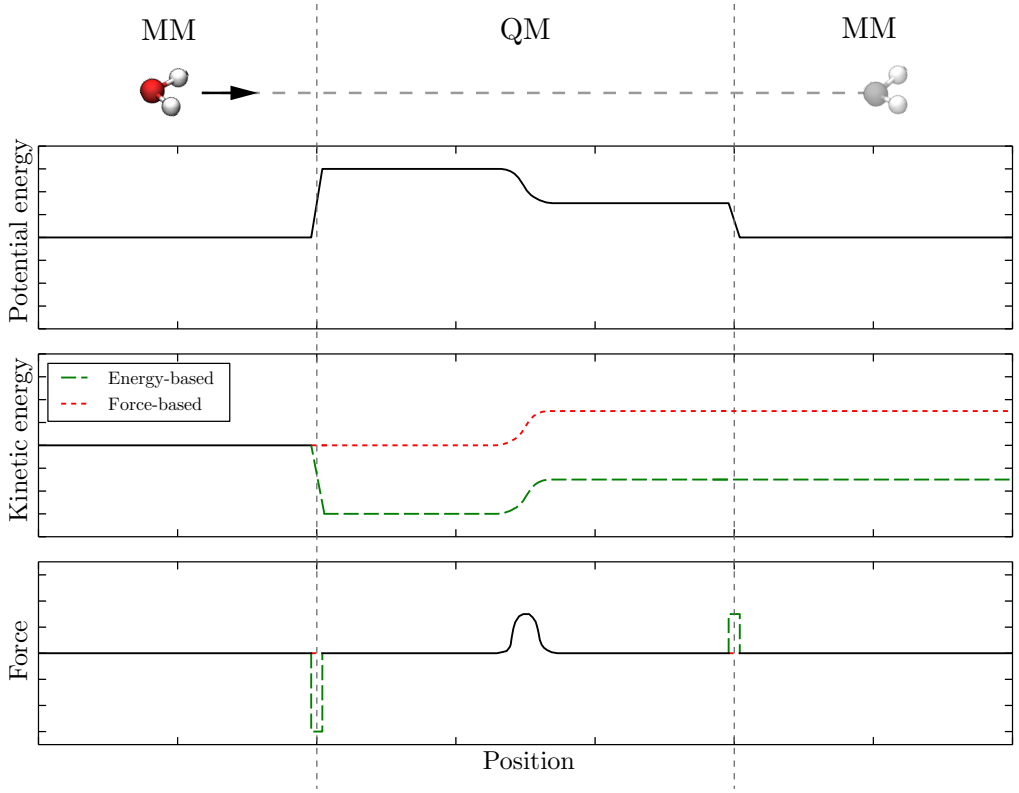


Figure 3.2: Illustration of differences between the adaptive methods when a molecule is moving from left to right and passing a QM region. In the energy-based methods, a fictitious force is introduced at the QM-MM interface due to the mismatch between potential energy surfaces. In the QM region the molecule experiences an interaction and part of it's potential energy is converted to kinetic energy. In the force-based methods this leads to violation in energy conservation when the molecule leaves the QM region.

In the so-called adaptive QM/MM methods, the particles are free to move between the QM and MM regions. This removes the need for ad hoc decisions that are made if the system is statically partitioned. This additional freedom does however complicate the energetical behaviour of the system.

If energy conservation is forced in an adaptive scheme, the forces must be made conservative in the adaptive scheme by calculating them from the change in the potential energy surface: $\vec{F}_i = -\nabla_i V$. This means that when an atom switches description any mismatch that exists between the MM and QM potential energy surfaces will give rise to unphysical forces near the boundary. These forces can lead to an artificial net flux of atoms between the QM and MM regions [35]. This effect can be removed by aligning the potential energy surfaces, but there however is no general approach for doing this. There are multiple adaptive schemes of this conservative type, such as ONIOM-XS [36], ABRUPT [37] and SAP [38, 39].

The alternative option is that at each time step of the dynamics, forces on atoms in the two regions are computed separately using the different models and then combined. This coupling in terms of forces can avoid the chemical potential mismatch effect, at the cost of forgoing energy conservation. This can eventually lead to unstable molecular dynamics trajectories, as kinetic energy can continue to increase/decrease near the boundary. This can however be avoided by using adaptive thermostats [35], which are briefly introduced in section 3.3.3.2. There are multiple of these force-based adaptive methods, such as adbf [35], DAS [40] and the Hot Spot method [41]. For our simulations the adbf (adaptive buffered-force) method was chosen and it is next described in more detail.

3.3.3.1 Adaptive Buffered Force QM/MM

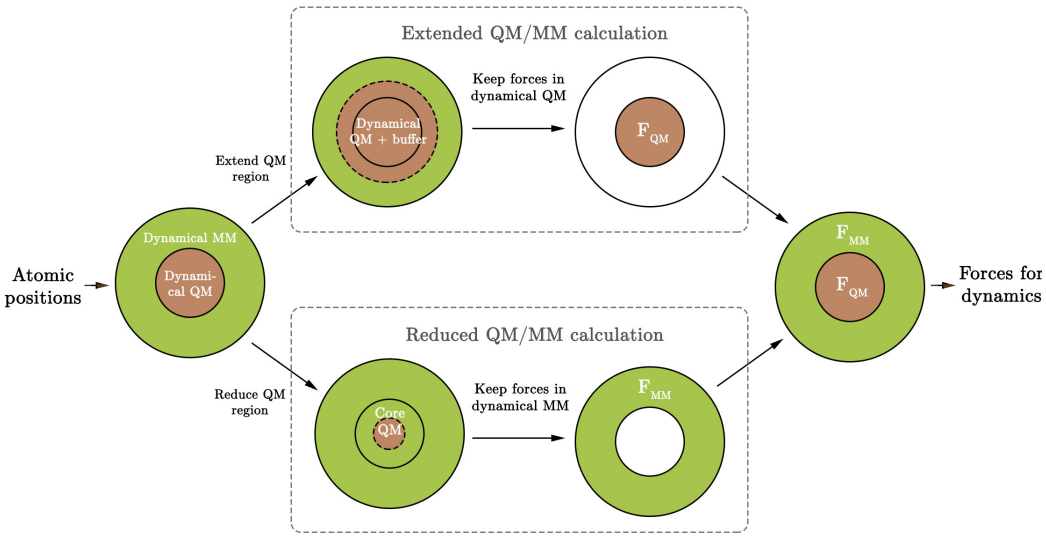


Figure 3.3: Flowchart for an adaptive buffered force QM/MM calculation.

In the adbf scheme the atoms are first divided into two groups: dynamical QM atoms which move according to QM forces and dynamical MM atoms which move according to the MM forces. The forces are however not calculated by using this same division, but are obtained by combining two QM/MM force calculations with a different partitioning. The flowchart describing the force calculation is shown in figure 3.3.

The forces for the dynamical QM region are calculated with an enlarged QM region ("extended QM/MM calculation") to obtain more accurate forces for atoms in the dynamical QM region. Using an extended QM region reduces the force error felt by the adaptive QM atoms near the boundary when compared to a fully QM force calculation for the same atomic positions. This extended QM region is constructed by adding a buffer region around the dynamical QM region. The buffer region size can be determined from the convergence of force errors in the dynamical QM region as a function of buffer region size.

The forces for the dynamical MM region are calculated with a second QM/MM calculation, typically now with a smaller QM region ("reduced QM/MM calculation"). By using a smaller QM region, the behaviour of the classical model can be obtained in the dynamical MM region more accurately and the computational time is reduced. If all the force field parameters are available, this calculation can be made even entirely classically. If for

example classical charge parameters are however not available for part of the dynamical QM atoms, these charges and the interaction can be calculated by using a suitably smaller QM region and electrical embedding.

Adaptivity is achieved by defining distance cutoffs to select atoms for the various regions that are dynamically evaluated at each time step during the simulation. To reduce the frequency of switching between regions for atoms that are close to the boundary, hysteresis is applied to all these distance cutoffs, so an atom has to come closer than some inner radius r_{in} to become incorporated into a region, but must move farther than a larger, outer radius r_{out} to be removed from the region. First, the core region is created by combining a fixed list and nearby atoms, based on a cutoff distance, r_{core}, from the atoms in the fixed list. Next, the dynamical QM region is defined as the union of the core region, another (optional) fixed list and atoms within a cutoff distance, r_{qm}, of core region atoms. Finally the buffer region is defined as the union of yet another optional fixed list and atoms within a cutoff distance, r_{buffer}, from atoms in the dynamical QM region. An example of such partitioning is shown in figure 3.4

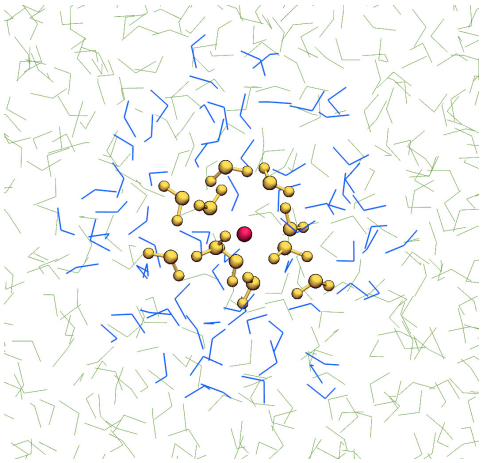


Figure 3.4: An example of system partitioning in the adbf scheme. The adaptive QM atoms moving according to the quantum forces are drawn with the ball-stick representation and adaptive MM atoms moving according to the classical forces are drawn as lines. Only one atom (red), is fixed to belong to the core QM region. The dynamical QM atoms (yellow) are determined by their distance from this core atom. The buffer QM atoms (blue) are determined by their distance from the adaptive QM atoms, and the rest of the system (green) is always classical.

When an atom is switched from the dynamical QM region to the dynam-

ical MM region or vice versa, the force it experiences will change instantly. A transition region could be used to smooth this transition, but the abrupt switching does not seem to significantly affect the accuracy of average structures and free energy profiles [35].

The non-conservative nature of the adbf scheme has two direct consequences. First, the lack of a total energy prevents the use of some free energy calculation methods, although potential of mean force methods, which require only forces and trajectories, can still be applied [35]. Second, because the forces are not the derivatives of any energy function, the dynamics are not conservative. This deviation from energy conservation requires the use of an appropriate thermostat, described in the next chapter. Additionally by calculating the forces with two different setups (the reduced and extended QM calculation) the force between atoms may not be equal thus violating Newton's third law and breaking the conservation of momentum. This deviation from linear momentum conservation can however be fixed exactly by adding a small correction force to some or all atoms.

3.3.3.2 Adaptive Thermostats

Because the atoms are allowed to change description, the energy of the system is time-dependent and thus is not conserved. This typically ends up generating heat near the boundary of the two regions [42]. In the case of artificical heating, the common thermostat methods such as NHC, NHL, or Langevin methods do not reproduce the intended target temperature and the distributions obtained are modified, often in a complicated way which does not facilitate recovery of correct thermodynamic averages [43].

The family of adaptive thermostats [43] have been shown to adapt to cancel average cooling or heating effects and sample the correct distribution in a fluctuating non-equilibrium environment, without a priori knowledge of the anomaly. Especially the Adaptive-Langevin thermostat can provide the strong internal equilibration property needed to control the distribution when the heating is rapid and non-uniform [43]. These properties make it ideal for QM/MM simulations with particle exchange between regions.

The used thermostat should also be "massive", meaning that a separate thermostat variable is coupled to each degree of freedom, rather than a single variable coupling to the total kinetic energy. This ensures more uniform temperature control as in the non-conservative force-mixing simulations heat anomalies might be generated near the QM-MM interface and the amount that needs to be dissipated therefore varies in space.

Chapter 4

Python interface for CP2K

The simulations in this work are all performed with the CP2K [29] open source simulation package. CP2K is a program to perform atomistic and molecular simulations of solid state, liquid, molecular, and biological systems. It provides a general framework for multiple different methods e.g. DFT with the Quickstep module [44] and classical pair and many-body potentials with the FIST module. These methods can be also then combined in various QM/MM setups.

Traditionally CP2K simulations are defined by creating static input files, with a xml-like structure. These files are then read by the CP2K executable. This is a very flexible, but often error prone and cumbersome way to define simulation setups. During this work an object-oriented scripting interface called PYCP2K was developed for CP2K.

PYCP2K allows the user to create and run entire CP2K simulations with python scripts. The simulation parameters can be dynamically created and altered thanks to an object tree that replaces the traditional CP2K inputs and other scripting tools like bash shell scripts. An example script is given in listing 4.1

The benefits of using PYCP2K over using the traditional CP2K input files:

- Intuitive notation with autocompletion (if supported by your IDE) for creating the input
- Define, run and analyze simulations with python scripts
- Fully modular simulation setup by e.g. using functions to define parts of the input
- Use python's control structures and arithmetic tools to define and manipulate input

- Structure creation and loading with ASE [45]
- Easily extract results from output file with predefined functions or with custom regular expressions
- Ability to build highly complex simulation frameworks

```
#! /usr/bin/env python
# -*- coding: utf-8 -*-

from pycp2k.cp2k import CP2K
from ase.lattice.cubic import Diamond

#================== Create the structure with ASE =====================
lattice = Diamond(directions=[[1, 0, 0], [0, 1, 0], [0, 0, 1]],
                  symbol='Si',
                  latticeconstant=5.430697500,
                  size=(1, 1, 1))

#============ Define and setup the calculator object =================
calc = CP2K()
calc.working_directory = "/home/lauri"
calc.project_name = "si_bulk"
calc.mpi_n_processes = 2

#=============== Define shortcuts for easy access ====================
CP2K_INPUT = calc.CP2K_INPUT
GLOBAL = CP2K_INPUT.GLOBAL
FORCE_EVAL = CP2K_INPUT.FORCE_EVAL_add()
SUBSYS = FORCE_EVAL.SUBSYS
DFT = FORCE_EVAL.DFT
SCF = DFT.SCF

#================== Write the simulation input ========================
GLOBAL.Run_type = "ENERGY_FORCE"
GLOBAL.Print_level = "LOW"
FORCE_EVAL.Method = "Quickstep"
FORCE_EVAL.PRINT.FORCES.Section_parameters = "ON"
DFT.Basis_set_file_name = "BASIS_SET"
DFT.Potential_file_name = "GTH_POTENTIALS"
DFT.QS.Eps_default = 1.0E-10
DFT.MGRID.Ngrids = 4
DFT.MGRID.Cutoff = 300
DFT.MGRID.Rel_cutoff = 60
DFT.XC.XC_FUNCTIONAL.Section_parameters = "PADE"
SCF.Scf_guess = "ATOMIC"
SCF.Eps_scf = 1.0E-7
SCF.Max_scf = 300
SCF.DIAGONALIZATION.Section_parameters = "ON"
SCF.DIAGONALIZATION.Algorithm = "STANDARD"
SCF.MIXING.Section_parameters = "T"
SCF.MIXING.Method = "BROYDEN_MIXING"
SCF.MIXING.Alpha = 0.4
SCF.MIXING.Nbroyden = 8
KIND = SUBSYS.KIND_add("Si")
KIND.Basis_set = "DZVP-GTH-PADE"
KIND.Potential = "GTH-PADE-q4"
calc.create_cell(SUBSYS, lattice)
calc.create_coord(SUBSYS, lattice)
```

```
#=================== Run the simulation ========================
calc.run()

#================ Gather and analyze results ==================
print calc.get_potential_energy()
print calc.get_forces()
print calc.get_output_value(r"SCF run converged in\s*(\d*)\s*steps")
```

Listing 4.1: An example of a simple PYCP2K script.

PYCP2K is freely available at `https://github.com/SINGROUP/pycp2k` and is under the GPL license.

Chapter 5

Methods

In this chapter the simulation setup and methods for analysis are described in detail. First a quick overview of the simulation setup is presented, and the choices for this setup are argumented. More detailed information about preparing and running the simulation setup is given in the end of this chapter.

5.1 Overview of Simulation Setup

Overview of the simulated system:

- 1000 water molecules, 1 ion
- A cubic simulation cell of $31.07\,\text{Å} \times 31.07\,\text{Å} \times 31.07\,\text{Å}$ resulting in the correct experimental water density
- Periodic boundary conditions to simulate bulk system
- Timestep of 0.5 fs

An overview of the classical force field setup:

- SPC/Fw water model [46]
- Lennard-Jones interaction between ion and oxygen in water [47], parameters optimized for SPC water model were used
- Ewald Summation for electrostatics
- Nose-Hoover chain thermostat when running pure MM dynamics

An overview of the DFT setup:

- BLYP functional [48, 49]
- GTH basis set [50]
- GTH pseudopotentials optimized for BLYP [51]
- Auxiliary plane-wave basis cutoff of 300 Ry and 450 Ry for systems with K and Na ions respectively.

- Grimme's DFT-D2 van Der Waals correction [52]

An overview of the QM/MM setup:

- Adaptive Langevin thermostat [53]
- Adaptive buffered force mixing [35]
  - Core QM region includes only the ion.
  - Dynamic QM radius of 3.0–3.4Å for Na, 3.4–3.8Å for K
  - Buffer QM radius of 3.2–3.6Å for Na, 3.3–3.7Å for K
- Electrical embedding with GEEP [30, 31], 12 Gaussians used. The cutoff radii r_c of 0.44 Å and 1.2 Å respectively for classical H and O atoms are taken from a previous simulation [30].

5.2 Choosing the Simulation Setup

In the QM/MM simulation the water molecules are allowed freely to diffuse between the different regions. This requires the use of a flexible classical water model. The SPC/Fw model was chosen due to the fact that it can well reproduce many of the thermodynamical, dynamical, and structural properties of water [46]. Due to the small differences in the behaviour of different water models, for each water model there exist optimized parametrizations for the ion-water Lennard–Jones interactions. Because of the low availability of these parameters for flexible water models, parameters optimized for the SPC water model were used instead. The validity of this approximation was ensured by comparing how similar the radial distribution function is compared to two other widely used classical force fields with optimized Lennard-Jones parameters. Checking the RDF should be enough as the Lennard-Jones interaction is radially symmetric.

As the exchange–correlation functional we use BLYP. It provides a suitable compromise between computational time and accuracy in describing water, and has been shown to better reproduce the oxygen-oxygen RDF in pure water than e.g. the PBE functional [54]. A DFT-D2 van Der Waals correction was included in the QM description as the vdW-corrected BLYP simulations give rise to highly mobile water whose softened structure is closer to experimental data than the one predicted by the bare BLYP functional [55]. At the same time this correction makes the QM and MM descriptions match better.

The convergence of total energy and behaviour of computational time in the system was tested against certain key simulation parameters. This allows one to determine a suitable compromise between accuracy and computational

time. The tested parameters include the G-space cutoff for the auxiliary plane-wave basis, the number of Gaussians in the GEEP expansion and the QM cell margin. Details can be found in appendix A. The QM cell margin was tested because the QM system has to be first calculated as an isolated system, by introducing a sufficient margin of empty space and using electrostatic decoupling with a multipole expansion [56]. This has to be done as the QM subsystem is smaller than the unit cell and thus the QM cell can't be simply periodically repeated in the calculations. This isolated system can be however recoupled to the periodic images with the correct periodicity of the unit cell by using the multipole charges.

The adbf implementation in CP2K doesn't support an empty core region. This would enable the second, reduced calculation to be made entirely classically. To best reproduce the valid MM behaviour in the dynamic MM region, the core region was instead made minimal and only includes the ion.

The dynamic QM radius determines what water molecules are dynamically included in the dynamic QM region. In our simulation only the ion and it's first hydration shell are intended to be modeled with the QM description. As an approximation for the dynamic radius we use the result from the classical simulation, i.e. the radius corresponding to the minimum between the first and second hydration shells in the radial distribution between the ion and the oxygens in the water. To allow small growth in the position of the first hydration shell in QM/MM, an additional 0.3 Å was added to the classical results of 3.2 Å and 3.6 Å for Na and K respectively. The hysteresis in the radius is included by introducing a transition region of 0.2 Å in both directions. From the previous computational results introduced in chapter 2.1.2, we can see that the position of the first hydration shell in the QM/MM or QM models should not too far away from the corresponding MM result, and this should be a valid choice.

The buffer radius in the adbf scheme determines the size of the extended quantum region. By increasing the buffer size, the dynamic QM molecules will transition from feeling the classical forces to feeling entirely QM forces from the atoms outside the dynamic region. To determine a good buffer radius in the adbf scheme, the average absolute force on the dynamic QM water molecules were calculated in five different snapshots from the classical simulation. From the convergence of these absolute forces against the buffer radius, a good approximation for the buffer size can be made assuming that the classical region in the QM/MM dynamics will have nearly the same behaviour as in the MM dynamics. The results can be found in appendix A. A buffer value of 4.2 corresponding to average absolute force error around 0.1 eV/Å was selected.

In order to roughly inspect the differences between the QM and MM

descriptions a spectral analysis comparison was performed on an isolated water dimer. Big differences in the vibrational modes correspond to bigger interfacial errors at the QM/MM border and weaker energetical coupling between the regions. Although these differences are always present and the use of an adaptive thermostat and force mixing alleviate the side effects, it is informative to know how well the two different descriptions match.

The validity of the dynamics is checked by investigating the temperature and energy behaviour during the simulation. During the production phase the temperature and the energy of the system (excluding the thermostat energy) should oscillate near a constant value, without any drifting. Due to the exchange of water molecules between descriptions in the QM/MM simulation, it is to be expected that the QM/MM energy will have discontinuities. However, due to the use of an adaptive thermostat, these exchange events shouldn't produce any heat anomalies, which would be shown in the system temperature.

5.3 Comparison between QM/MM and MM

The validity of the QM/MM simulations is measured by comparing the results to previous experimental and computational results and also to a corresponding classical simulation that was run on the same system. The QM/MM simulations are run with both mechanical embedding and electrical embedding to investigate the differences in dynamical properties and in computational time. This allows one to identify whether the more elaborate and computationally expensive electrical embedding is necessary in simple ion-water solutions. A fully quantum mechanical simulation was not performed for comparison due to it's computational expensiveness and also due to the availability of previous computational results. A comparison between the computational times for one step of molecular dynamics for the MM, QM/MM and QM methods was also made.

The calculated structural and dynamical properties of the system include the radial distribution function between the ion and the oxygen atoms in the water, the hydration number of the first shell, the angular distribution of oxygen–ion–oxygen triplets inside the first solvation shell and the mean residence time of water molecules in the first hydration shell. With these values one can achieve a quantitative structural description of the first solvation shell that can be then compared between models. The minima and maxima in the radial distribution functions were determined with an optimal Gaussian fit. To smooth the radial distribution function and the O–ion-O probability, a Savitsky-Golay filter with third order polynomials was applied.

This smoothing increases the readability especially for the QM/MM simulations which have much more noise compared to the MM simulation due to shorter simulation times.

5.4 Simulation Procedure

Here is provided a more detailed account of how the simulated system was constructed and prepared before running the dynamics and how the data was gathered from the production runs.

5.4.1 Initialization

In order to bring the system to a reasonable initial condition before starting the production run, the system undergoes optimization, heating and equilibration. These steps are common for both MM and QM/MM runs. These steps are entirely classical, as we are only searching for a reasonable starting configuration for the system.

The simulated solution consists of a one solute ion in a pure water solvent consisting of $10 \times 10 \times 10 = 1000$ water molecules. A cubic simulation cell of $31.07\,\text{Å} \times 31.07\,\text{Å} \times 31.07\,\text{Å}$ was selected to reproduce the experimental water density of $0.997\,05\,\text{kg/dm}^3$ [57] at 25 °C and 100 kPa. Periodic boundary conditions were employed to simulate a bulk solution. This setup should be large enough to remove most of the PBC artifacts and allow the ion to interact with multiple water molecules.

Initially the water molecules were placed in a uniformly spaced grid and the ion was placed directly in the middle of the simulation cell. This geometry was then optimized with the conjugate gradient method. This is done to remove any bigger tensions from the initial configuration which might inadvertently affect the dynamics. Highly energetic initial conditions will lead to slower equilibration and might also produce anomalous behaviour for individual atoms that would be hard to detect during dynamics.

The system was then heated to the temperature of 298.15 K with a Nose-Hoover chain thermostat. Three chains were employed with a time constant of 100 fs for fast equilibration. After heating the positions and velocities of the atoms were stored for starting the equilibration.

Before starting the production runs, the system was classically equilibrated, now starting much closer to the desired temperature. This equilibration was done with the same Nose-Hoover thermostat, but now with a time-constant of 200 fs. The equilibration was performed for 5 ps before starting the production runs.

5.4.2 Production Phase

The MM production runs were started after the classical equilibration phase. For this production run the same Nose-Hoover time constant of 200 fs was used. The production runs were performed for 95 ps.

The abrupt transition to a QM/MM simulation will introduce a discontinuity in temperature and energy, and thus a reequilibration of 0.5 ps with the QM/MM setup is needed. The thermostat was changed to an adaptive Langevin thermostat. The time constant for the Langevin part of the thermostat was set to 70 fs and for Nose-Hoover part to 140 fs. These shorter time constants allow faster equilibration that is needed due to much shorter simulation times. The QM/MM production runs were performed for 2.5 ps and 7 ps for the Na^+ and K^+ ions respectively. The shorter simulation time for sodium is due to twice larger computational time.

Chapter 6

Results

6.1 Validating the MM model

Figure 6.1 shows the radial distribution function between the ion and the oxygen of the water for three classical models. The characteristics of the first hydration layer for these models are given in table 6.1.

6.2 Spectral Density Comparison

The spectral density of an isolated water dimer was calculated separately with the MM and QM model. The dimer system is the smallest system that can reveal the intra-molecular vibrations as well as the low frequency inter-molecular vibrations between two water molecules. The simulations were carried for 40 ps with the setups described in section 5.1, but now with a cutoff of 340 Ry in the QM simulation.

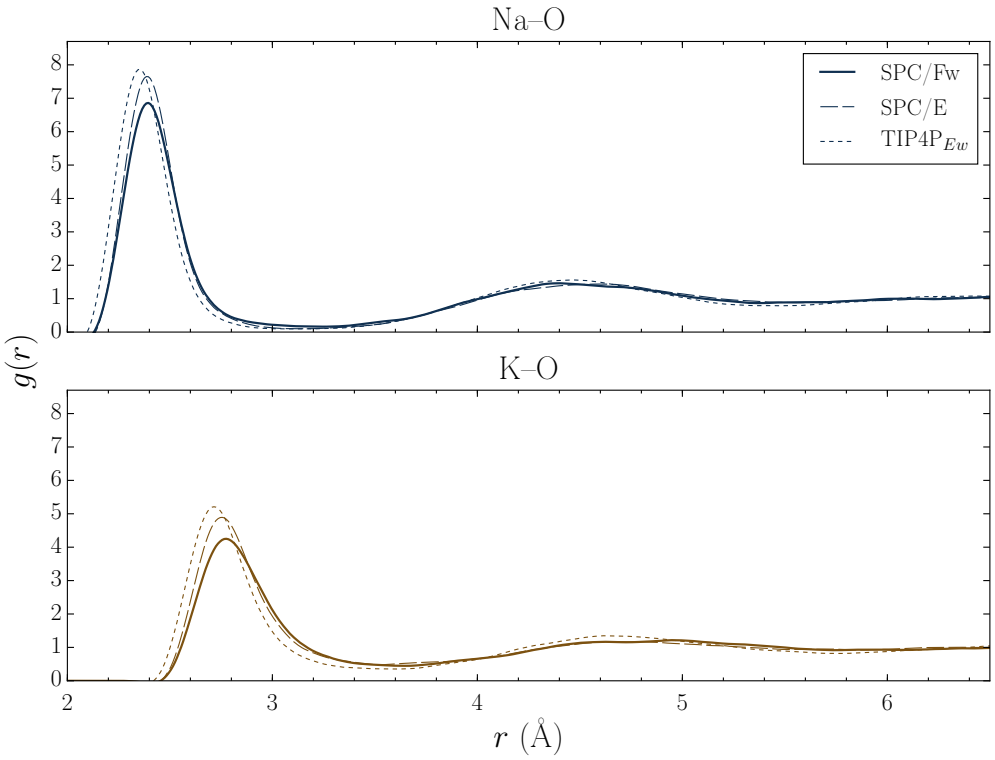


Figure 6.1: The radial distribution function between sodium and oxygen (upper panel) and potassium and oxygen (lower panel) for the three different classical water models.

Table 6.1: Comparison of the first hydration layer properties for the Na^+ and K^+ ions between different classical water models. r_{max1} refers to the first maximum in the ion–oxygen radial distribution function. N_1 refers to the hydration number of the first hydration shell.

Ion	Method	r_{max1}	N_1
Na^+	SPC/Fw	2.38	5.6
	SPC/E	2.37	5.8
	TIP4P$_{EW}$	2.34	5.8
K^+	SPC/Fw	2.76	6.9
	SPC/E	2.74	7.4
	TIP4P$_{EW}$	2.70	7.0

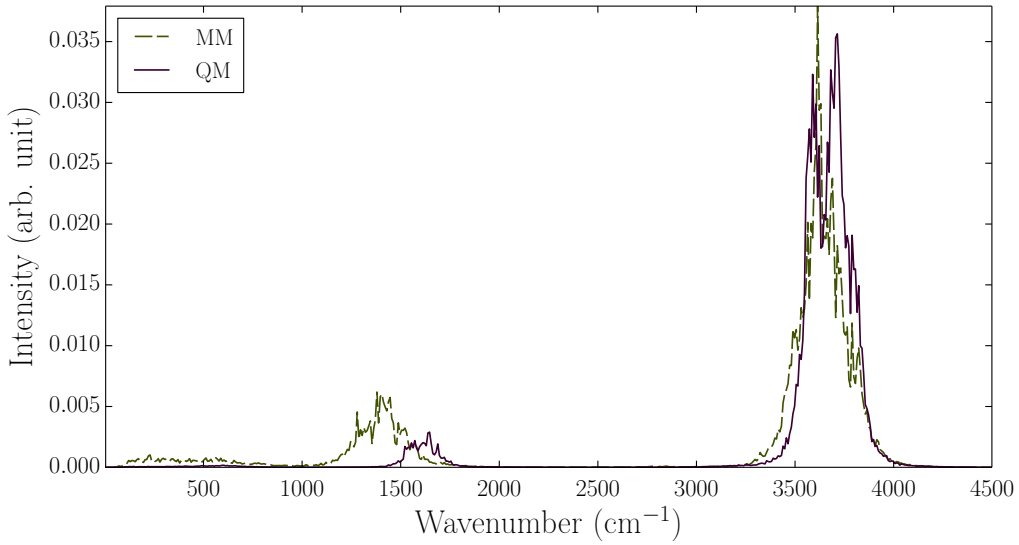


Figure 6.2: The spectral density of different vibrations in an isolated water dimer for both QM and MM descriptions.

6.3 Energy and Temperature Behaviour during MD

Figures 6.3 and 6.4 show the temperature and the energy for the MM and QM/MM simulations with Na^+ ion. The exact same characteristics are seen with the K^+ ion.

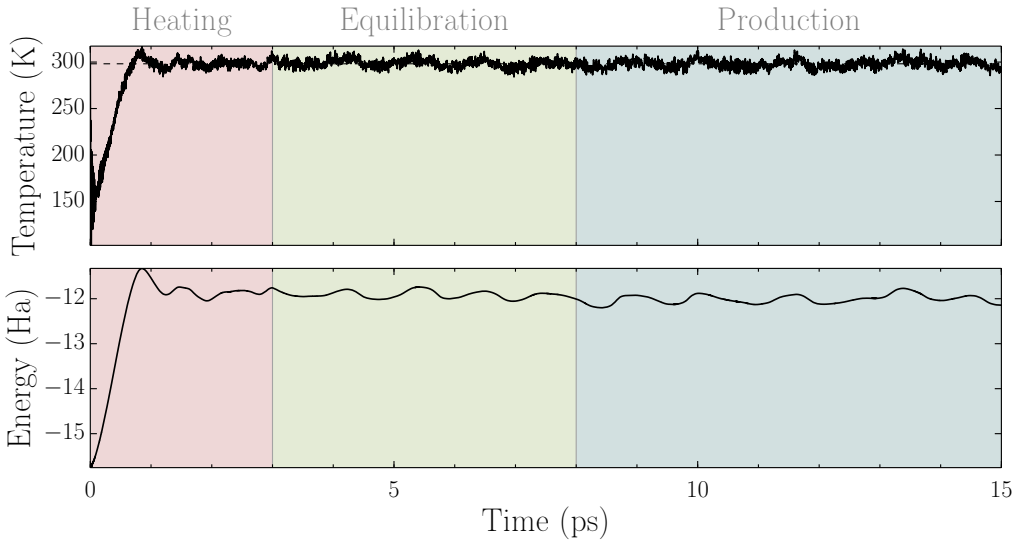


Figure 6.3: The temperature and energy of the classical simulation with Na^+ ion. The energy consists of the combined kinetic and potential energy of the system, excluding the thermostat energy.

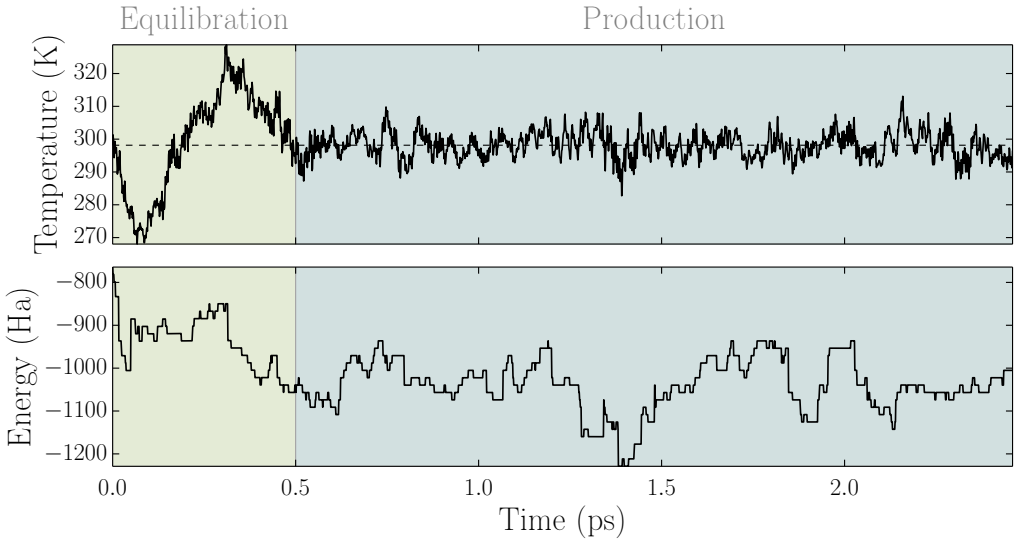


Figure 6.4: The temperature and energy of the QM/MM simulation with Na^+ ion. Part of the production run has been cut out. The energy consists of the combined kinetic and potential energy of the extended QM/MM calculation, excluding the thermostat energy.

The temperature around the area where atoms switch from MM forces to QM forces or vice versa is of special interest due to the non-conservative forces that might heat or cool the system. To investigate if such temperature anomalies were locally present, the temperature of atoms within spherical shells around the ion were investigated. The results are shown in figure 6.5.

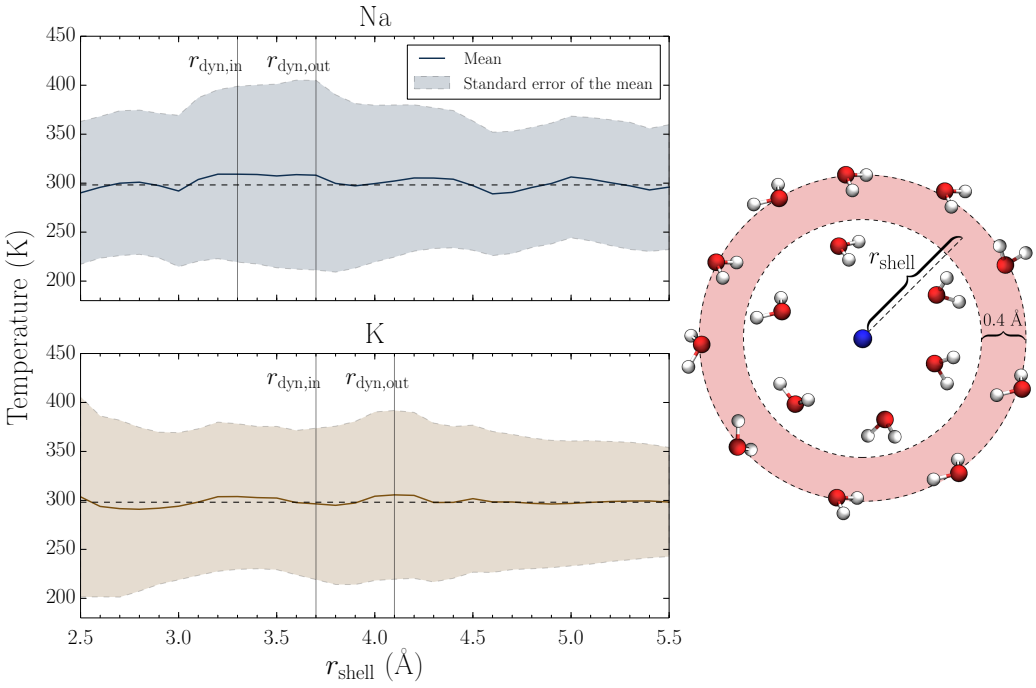


Figure 6.5: The local temperature around the ions in the QM/MM simulations with electrical embedding. The average temperature and it's standard error were calculated from the production run. The temperature was calculated for atoms within a spherical shell centered on the ion. r_{shell} is the distance from the ion to the halfway between the inner and outer radius of the shell, shell being 0.4 Å thick.

6.4 Comparison between QM/MM and MM

The results from QM/MM simulations with electrical embedding and with mechanical embedding are compared to each other and to the classical reference simulations. The radial distributions are found in figure 6.6, angular distribution within the first hydration shell in figure 6.7 and a comparison of the first hydration layer characteristics in table 6.2. A comparison of the computational time for different methods is shown in figure 6.8.

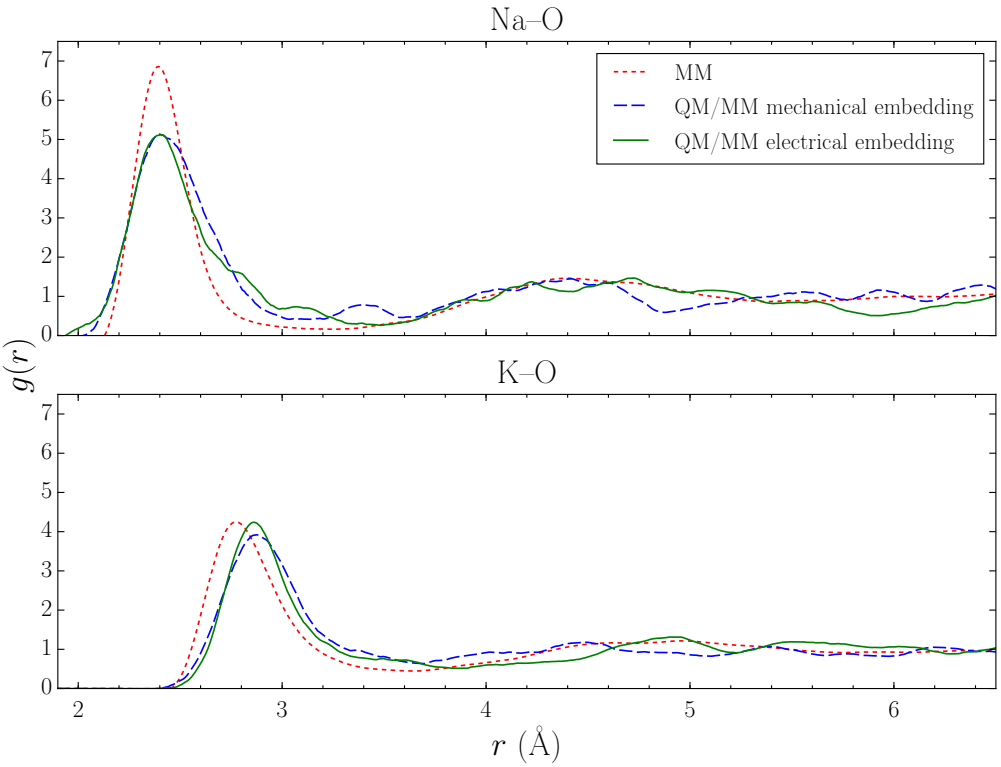


Figure 6.6: The ion-oxygen radial distribution function from the MM and QM/MM simulations.

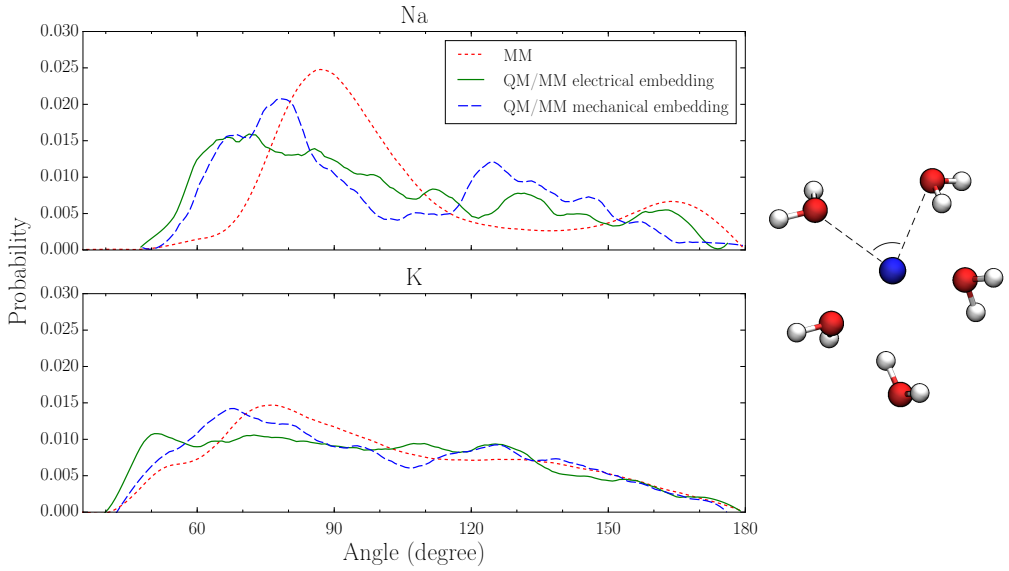


Figure 6.7: The distribution of the oxygen–ion–oxygen angle within the first hydration layer for the MM and QM/MM simulations.

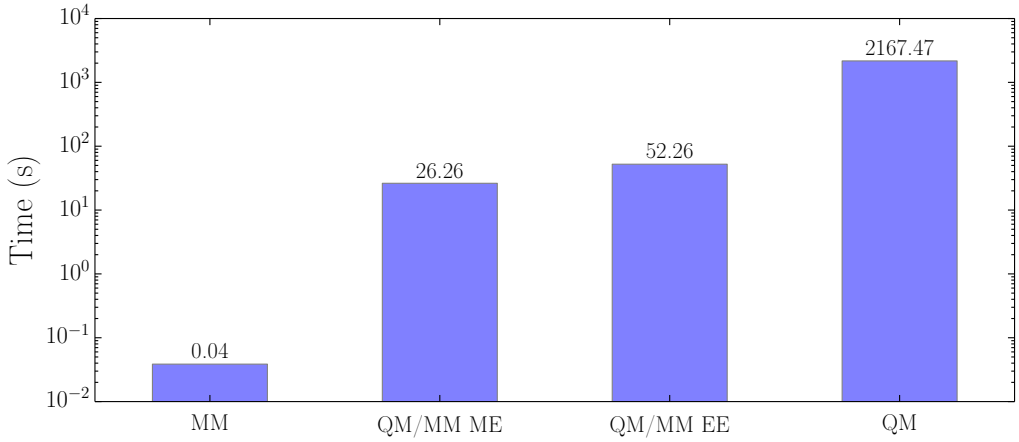


Figure 6.8: The time taken to perform one step of molecular dynamics for the potassium ion by each method. Calculations were made on identical machines with 60 processor cores in use.

Table 6.2: Comparison of first hydration shell properties of Na^+ and K^+ ions. r_{max1} refers to the first maximum in the ion–oxygen radial distribution function. N_1 refers to the coordination number of the first hydration shell. τ_0 and $\tau_{0.5}$ refer to the mean residence times with time constants of 0 and 0.5 ps respectively.

Ion	Method	r_{max1}	N_1	τ_0	$\tau_{0.5}$
Na^+	MM	2.38	5.6	1.92	3.84
	QM/MM ME	2.42	6.9	0.53	3.44
	QM/MM EE	2.39	7.1	1.01	2.82
K^+	MM	2.76	6.9	0.58	1.64
	QM/MM ME	2.88	8.1	0.51	1.87
	QM/MM EE	2.86	8.4	0.58	1.71

Chapter 7

Discussion and Conclusions

The use of non-optimized oxygen–ion Lennard-Jones parameters with the SPC/Fw model produces a rdf-distribution compatible with other optimized classical models. This result validates the use of these parameters to produce MM results of similar quality with the other classical models.

The comparison of spectral densities reveals a decent match between the different vibrational modes. Especially the highest frequencies stemming from the O–H oscillations match very well. The match in vibrational frequencies also indicates the match between forces in the two models. A better match in forces means that any possible negative effects from the force discontinuity and momentum conservation fix are reduced. The exchange of kinetic energy through the QM/MM border by vibrational coupling might also have a minor effect on the dynamics. This will also be controlled by the match in the vibrational modes.

The system temperature and energies behave as expected in the MM and QM/MM simulations. The energetical exchange between the atoms and the thermostat creates an oscillating base frequency in the energy, but no drift is observed. In the QM/MM simulations a higher frequency sharp oscillations can be seen which stem from the atoms switching between QM and MM regions. When examining the local temperature in a radial shell around the ion, no heat anomalies can be detected. This indicates that the massive adaptive thermostat can keep the energy fluctuations in order, or there is no significant heat generation/loss at this timescale in this particular setup.

From the radial distribution function it can be seen that the classical model seems to overstructure the water in the first hydration layer around sodium. With potassium the shape of the first hydration layer is very similar for all methods but the position of the first hydration layer is smaller in MM than in QM/MM. This can also be seen in the previous computational results. The order of $r_{\mathrm{max1}}^{\mathrm{MM}} < r_{\mathrm{max1}}^{\mathrm{QM/MM\ EE}} < r_{\mathrm{max1}}^{\mathrm{QM/MM\ EE}}$ can be observed from

table 6.2 for both ions, although the differences are small. No unphysical density anomalies can be seen at the border of the dynamical QM and MM regions, which is to be expected from the force-based adaptive scheme.

The classical angular distribution for sodium reveals a highly ordered octahedral structure with peaks near 90° and 180° at roughly 4/1 ratio. This structuring is not observed in the QM/MM simulations. The QM/MM simulations have more of the smaller angles present which correspond to the observed higher hydration number. Both QM/MM versions produce more uniform distributions than the classical version but the mechanical embedding seems to still produce two distinct peaks around 75° and 135°. The angular distribution for potassium is highly uniform with all methods. The higher hydration number can be seen from the presence of smaller angles than in the distribution for sodium.

The hydration number of potassium agrees well with the previous computational results. The hydration number obtained for sodium is higher than in any previous result, but this can be due to the inaccuracy in determining where the first hydration layer ends. Longer simulations might produce a better defined minimum between the first and second hydration layer.

The mean residence times for potassium are very similar for all the methods. For sodium the classical model produces the most stable binding, which is in line with having the most structured radial and angular distribution functions. Between the two embedding schemes, an order $\tau_0^{ME} < \tau_0^{EE}$ and $\tau_{0.5}^{ME} > \tau_{0.5}^{EE}$ can be seen for both ions. For all methods the mean residence times τ_0 and $\tau_{0.5}$ for sodium are always higher than the corresponding times for potassium. This is in line with the previous computational results.

The comparison of computational times for one step of molecular dynamics reveals large differences as expected. From MM to QM/MM the computational time has increased roughly by a factor of 10^3, and from QM/MM to QM the time would increase by roughly 10^2. The electrical embedding with GEEP will computationally cost roughly twice as much compared to mechanical embedding with point charges. These values are of course highly specific to the simulation setup and for example the use of an extended QM region in adbf increases the computational time compared to a more traditional QM/MM partitioning.

The QM/MM model was able to produce realistic simulations that largely agree with previous computational results. A full QM simulation for a similar system could be used to better validate the QM description of the first hydration shell in the QM/MM model and to investigate how smooth the transition from QM to MM is. For sodium the QM/MM model can be distinguished from the MM model due to observable differences in all the measured properties of the first hydration shell. These differences seem to

arise from energetically stronger structuring of the first hydration shell in the MM model. For potassium the MM model gives almost equal description as QM/MM. The differences between mechanical and electrical embedding are minimal in this system. Due to the very homogeneously distributed charges in the classical region, there are no strongly oriented or high electrical fields that would more strongly polarize the electron density in electrical embedding. In systems with larger localised charges or uniform electric fields the difference could become crucial. Due to the advantage in computational time for the mechanical embedding, it can be safely recommended for these types of systems.

Bibliography

[1] Veronika Brazdova and David R. Bowler. *Atomistic Computer Simulations: A Practical Guide*. Wiley-VCH, 2013. ISBN 3527410694.

[2] The Royal Swedish Academy of Sciences. Scientific background on the nobel prize in chemistry 2013: Development of multiscale models for complex chemical systems, 2013.

[3] Philippe Hünenberger. *Single-ion solvation : experimental and theoretical approaches to elusive thermodynamic quantities*. RSC Pub, Cambridge, 2011. ISBN 978-1847551870.

[4] Arthur Guyton. *Textbook of medical physiology*. Elsevier Saunders, Philadelphia, 2006. ISBN 0-7216-0240-1.

[5] Johan Mähler and Ingmar Persson. A study of the hydration of the alkali metal ions in aqueous solution. *Inorganic Chemistry*, 51(1):425–438, 2012. doi: 10.1021/ic2018693.

[6] Yizhak Marcus. Effect of ions on the structure of water: structure making and breaking. *Chemical reviews*, 109(3):1346–1370, 2009.

[7] G. Palinkas, T. Radnal, and F. Hajdu. Ion-solvent and solvent-solvent interactions. x-ray study of aqueous alkali chloride solutions. *Z. Naturforschung. A*, 35:107 – 114, 1980.

[8] N.T. Skipper and G. W. Neilson. X-ray and neutron diffraction studies on concentrated aqueous solutions of sodium nitrate and silver nitrate. *Journal of Physics: Condensed Matter*, 1(26):4141, 1989. URL `http://stacks.iop.org/0953-8984/1/i=26/a=010`.

[9] S. Ansell, A.C. Barnes, P.E. Mason, G.W. Neilson, and S. Ramos. X-ray and neutron scattering studies of the hydration structure of alkali ions in concentrated aqueous solutions. *Biophysical Chemistry*, 124(3):171 –

179, 2006. ISSN 0301-4622. doi: http://dx.doi.org/10.1016/j.bpc.2006.04.018.

[10] R. Mancinelli, A. Botti, F. Bruni, M. A. Ricci, and A. K. Soper. Hydration of sodium, potassium, and chloride ions in solution and the concept of structure maker/breaker. *The Journal of Physical Chemistry B*, 111 (48):13570–13577, 2007. doi: 10.1021/jp075913v.

[11] Johan Mähler and Ingmar Persson. A study of the hydration of the alkali metal ions in aqueous solution. *Inorganic Chemistry*, 51(1):425–438, 2012. doi: 10.1021/ic2018693.

[12] In Suk Joung and Thomas E. Cheatham. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *The Journal of Physical Chemistry B*, 112(30):9020–9041, 2008. doi: 10.1021/jp8001614.

[13] Song Hi Lee and Jayendran C. Rasaiah. Molecular dynamics simulation of ionic mobility. i. alkali metal cations in water at 25° C. *The Journal of Chemical Physics*, 101(8):6964–6974, 1994. doi: http://dx.doi.org/10.1063/1.468323.

[14] S. Koneshan, Jayendran C. Rasaiah, R. M. Lynden-Bell, and S. H. Lee. `Solvent Structure, Dynamics, and Ion Mobility in Aqueous Solutions at` 25° C. *The Journal of Physical Chemistry B*, 102(21):4193–4204, 1998. doi: 10.1021/jp980642x.

[15] Anan Tongraar, Klaus R. Liedl, and Bernd M. Rode. Born–oppenheimer ab initio qm/mm dynamics simulations of na+ and k+ in water: From structure making to structure breaking effects. *The Journal of Physical Chemistry A*, 102(50):10340–10347, 1998. doi: 10.1021/jp982270y.

[16] Anan Tongraar and Bernd M. Rode. Dynamical properties of water molecules in the hydration shells of na^+ and k^+: ab initio qm/mm molecular dynamics simulations. *Chemical Physics Letters*, 385(5–6): 378 – 383, 2004. doi: http://dx.doi.org/10.1016/j.cplett.2004.01.010.

[17] S. Sikander Azam, Zaheer ul Haq, and M. Qaiser Fatmi. Classical and qm/mm md simulations of sodium(i) and potassium(i) ions in aqueous solution. *Journal of Molecular Liquids*, 153(2–3):95 – 100, 2010. ISSN 0167-7322. doi: http://dx.doi.org/10.1016/j.molliq.2010.01.005. URL `http://www.sciencedirect.com/science/article/pii/S0167732210000152`.

[18] Christopher N. Rowley and Benoît Roux. The solvation structure of na^+ and k^+ in liquid water determined from high level ab initio molecular dynamics simulations. *Journal of Chemical Theory and Computation*, 8 (10):3526–3535, 2012. doi: 10.1021/ct300091w.

[19] Susan B. Rempe and Lawrence R. Pratt. The hydration number of na^+ in liquid water. *Fluid Phase Equilibria*, 183–184(0):121 – 132, 2001. ISSN 0378-3812. doi: http://dx.doi.org/10.1016/S0378-3812(01)00426-5.

[20] Jody A. White, Eric Schwegler, Giulia Galli, and François Gygi. The solvation of na^+ in water: First-principles simulations. *The Journal of Chemical Physics*, 113(11):4668–4673, 2000.

[21] Supachai Wanprakhon, Anan Tongraar, and Teerakiat Kerdcharoen. Hydration structure and dynamics of k+ and ca2+ in aqueous solution: Comparison of conventional qm/mm and oniom-xs md simulations. *Chemical Physics Letters*, 517(4–6):171 – 175, 2011. ISSN 0009-2614. doi: http://dx.doi.org/10.1016/j.cplett.2011.10.048.

[22] Susan B. Rempe, D. Asthagiri, and Lawrence R. Pratt. Inner shell definition and absolute hydration free energy of k+(aq) on the basis of quasi-chemical theory and ab initio molecular dynamics. *Phys. Chem. Chem. Phys.*, 6:1966–1969, 2004. doi: 10.1039/B313756B.

[23] Hai Lin and Donald G. Truhlar. Qm/mm: what have we learned, where are we, and where do we go from here? *Theoretical Chemistry Accounts*, 117(2):185–199, 2007. ISSN 1432-881X. doi: 10.1007/s00214-006-0143-z. URL `http://dx.doi.org/10.1007/s00214-006-0143-z`.

[24] Dominik Marx and Hutter Jürg. *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*. Cambridge University Press, Cambridge, 2012. ISBN 9781107663534.

[25] Wolfram Koch and Max C. Holthausen. *A chemist's guide to density functional theory*. Wiley-VCH, Weinheim New York, 2001. ISBN 978-3-527-30372-4.

[26] David Griffiths. *Introduction to quantum mechanics*. Pearson Prentice Hall, Upper Saddle River, NJ, 2005. ISBN 0131911759.

[27] Noam Bernstein, James R Kermode, and Gabor Csanyi. Hybrid atomistic simulation methods for materials systems. *Reports on Progress in Physics*, 72(2):026501, 2009.

[28] Dirk Bakowies and Walter Thiel. Hybrid models for combined quantum mechanical and molecular mechanical approaches. *Journal of Physical Chemistry*, 100(25):10580–10594, 1996. ISSN 1432-881X. doi: 10.1021/jp9536514.

[29] Cp2k open source molecular dynamics. `http://www.cp2k.org/`. Accessed: 2015-04-23.

[30] Teodoro Laino, Fawzi Mohamed, Alessandro Laio, and Michele Parrinello. An efficient real space multigrid qm/mm electrostatic coupling. *Journal of Chemical Theory and Computation*, 1(6):1176–1184, 2005.

[31] Teodoro Laino, Fawzi Mohamed, Alessandro Laio, and Michele Parrinello. An efficient linear-scaling electrostatic coupling for treating periodic boundary conditions in qm/mm simulations. *Journal of Chemical Theory and Computation*, 2(5):1370–1378, 2006.

[32] Vincent Théry, Daniel Rinaldi, Jean-Louis Rivail, Bernard Maigret, and György G Ferenczy. Quantum mechanical computations on very large molecular systems: The local self-consistent field method. *Journal of computational chemistry*, 15(3):269–282, 1994.

[33] Jiali Gao, Patricia Amara, Cristobal Alhambra, and Martin J Field. A generalized hybrid orbital (gho) method for the treatment of boundary atoms in combined qm/mm calculations. *The Journal of Physical Chemistry A*, 102(24):4714–4721, 1998.

[34] Yingkai Zhang, Tai-Sung Lee, and Weitao Yang. A pseudobond approach to combining quantum mechanical and molecular mechanical methods. *The Journal of chemical physics*, 110(1):46–54, 1999.

[35] Letif Mones, Andrew Jones, Andreas W Götz, Teodoro Laino, Ross C Walker, Ben Leimkuhler, Gábor Csányi, and Noam Bernstein. The adaptive buffered force qm/mm method in the cp2k and amber software packages. *arXiv preprint arXiv:1409.5218*, 2014.

[36] Teerakiat Kerdcharoen and Keiji Morokuma. Oniom-xs: an extension of the oniom method for molecular simulation in condensed phase. *Chemical physics letters*, 355(3):257–262, 2002.

[37] Rosa E Bulo, Carine Michel, Paul Fleurat-Lessard, and Philippe Sautet. Multiscale modeling of chemistry in water: are we there yet? *Journal of Chemical Theory and Computation*, 9(12):5567–5577, 2013.

[38] Andreas Heyden, Hai Lin, and Donald G Truhlar. Adaptive partitioning in combined quantum mechanical and molecular mechanical calculations of potential energy functions for multiscale simulations. *The Journal of Physical Chemistry B*, 111(9):2231–2241, 2007.

[39] Andreas Heyden and Donald G Truhlar. Conservative algorithm for an adaptive change of resolution in mixed atomistic/coarse-grained multiscale simulations. *Journal of Chemical Theory and Computation*, 4(2): 217–221, 2008.

[40] Rosa E Bulo, Bernd Ensing, Jetze Sikkema, and Lucas Visscher. Toward a practical method for adaptive qm/mm simulations. *Journal of Chemical Theory and Computation*, 5(9):2212–2221, 2009.

[41] Teerakiat Kerdcharoen, Klaus R Liedl, and Bernd M Rode. A qm/mm simulation method applied to the solution of li+ in liquid ammonia. *Chemical physics*, 211(1):313–323, 1996.

[42] Noam Bernstein, Csilla Várnai, Ivan Solt, Steven A Winfield, Mike C Payne, István Simon, Mónika Fuxreiter, and Gábor Csányi. Qm/mm simulation of liquid water with an adaptive quantum region. *Physical Chemistry Chemical Physics*, 14(2):646–656, 2012.

[43] Andrew Jones and Ben Leimkuhler. Adaptive stochastic methods for sampling driven molecular systems. *The Journal of chemical physics*, 135(8):084125, 2011.

[44] Joost VandeVondele, Matthias Krack, Fawzi Mohamed, Michele Parrinello, Thomas Chassaing, and Jürg Hutter. Quickstep: Fast and accurate density functional calculations using a mixed gaussian and plane waves approach. *Computer Physics Communications*, 167(2):103–128, 2005.

[45] S. R. Bahn and K. W. Jacobsen. An object-oriented scripting interface to a legacy electronic structure code. *COMPUTING IN SCIENCE & ENGINEERING*, 4(3):56–66, 2002. ISSN 1521-9615. doi: 10.1109/5992.998641.

[46] Yujie Wu, Harald L Tepper, and Gregory A Voth. Flexible simple point-charge water model with improved liquid-state properties. *The Journal of chemical physics*, 124(2):024503, 2006.

[47] Maria M Reif and Philippe H Hünenberger. Computation of methodology-independent single-ion solvation properties from molecular simulations. iv. optimized lennard-jones interaction parameter sets for the alkali and halide ions in water. *The Journal of chemical physics*, 134(14):144104, 2011.

[48] A. D. Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A*, 38:3098–3100, Sep 1988. doi: 10.1103/PhysRevA.38.3098.

[49] Chengteh Lee, Weitao Yang, and Robert G Parr. Development of the colle-salvetti correlation-energy formula into a functional of the electron density. *Physical review B*, 37(2):785, 1988.

[50] Joost VandeVondele and Juerg Hutter. Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases. *The Journal of chemical physics*, 127(11):114105, 2007.

[51] M Krack. Pseudopotentials for h to kr optimized for gradient-corrected exchange-correlation functionals. *Theoretical Chemistry Accounts*, 114 (1-3):145–152, 2005.

[52] Stefan Grimme. Semiempirical gga-type density functional constructed with a long-range dispersion correction. *Journal of computational chemistry*, 27(15):1787–1799, 2006.

[53] Andrew Jones and Ben Leimkuhler. Adaptive stochastic methods for sampling driven molecular systems. *The Journal of chemical physics*, 135(8):084125, 2011.

[54] I-Chun Lin, Ari P. Seitsonen, Ivano Tavernelli, and Ursula Rothlisberger. Structure and dynamics of liquid water from ab initio molecular dynamics—comparison of blyp, pbe, and revpbe density functionals with and without van der waals corrections. *Journal of Chemical Theory and Computation*, 8(10):3902–3910, 2012. doi: 10.1021/ct3001848.

[55] I-Chun Lin, Ari P. Seitsonen, Maurício D. Coutinho-Neto, Ivano Tavernelli, and Ursula Rothlisberger. Importance of van der waals interactions in liquid water. *The Journal of Physical Chemistry B*, 113(4):1127–1131, 2009. doi: 10.1021/jp806376e.

[56] PE Blöchl. Electrostatic decoupling of periodic images of plane-wave-expanded densities and derived atomic point charges. *The Journal of chemical physics*, 103(17):7422–7428, 1995.

[57] W.M. Haynes. *CRC Handbook of Chemistry and Physics, 95th Edition*. CRC Handbook of Chemistry and Physics. Taylor & Francis, 2014. ISBN 9781482208672.

Appendix A

Convergence Tests

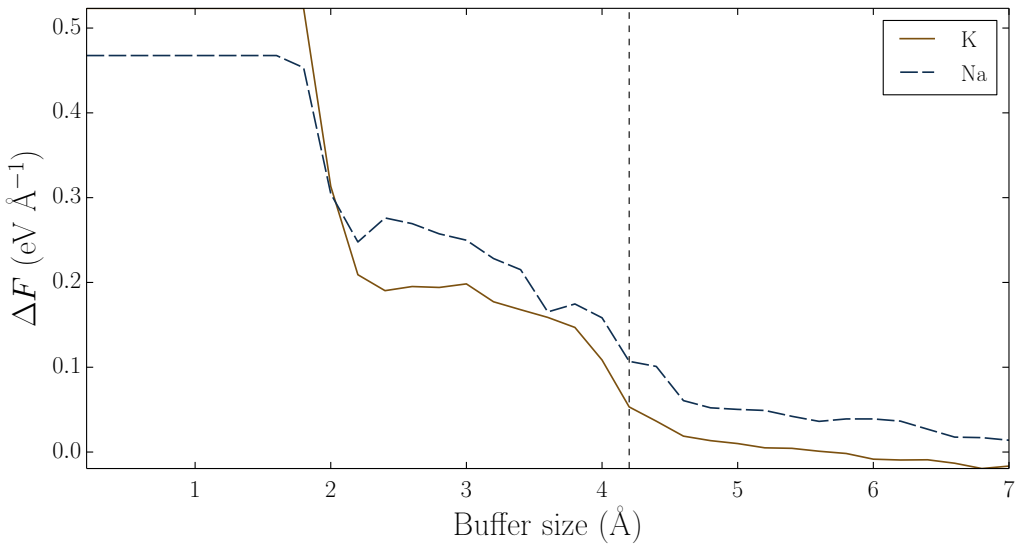


Figure A.1: The error in mean absolute force for the dynamic water molecules as a function of buffer size. The forces are calculated in five different snapshots taken from the MM simulation. The error is calculated relative to the full QM forces in the same setup. The value of 4.2 Å chosen for both ions is shown with a vertical line.

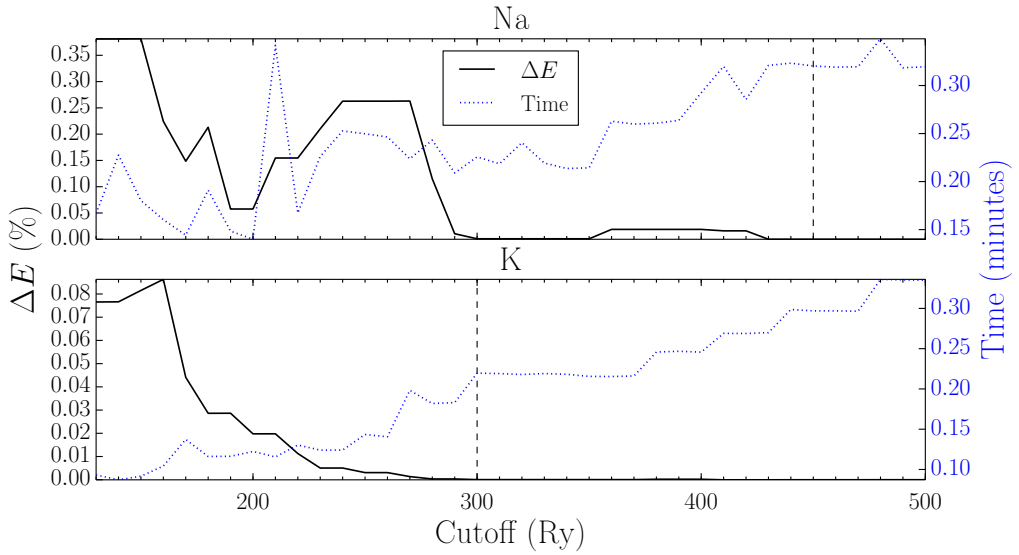


Figure A.2: The energy convergence and elapsed time for sodium and potassium with respect to the cutoff value for the auxiliary plane-wave basis used in CP2K. The energy is compared the last tested cutoff energy of 500 Ry. The vertical lines indicates the values 450Ry and 300Ry chosen for Na and K respectively. The test system consists of the respective ion surrounded by six closest water molecules where positions are extracted from classical MD.

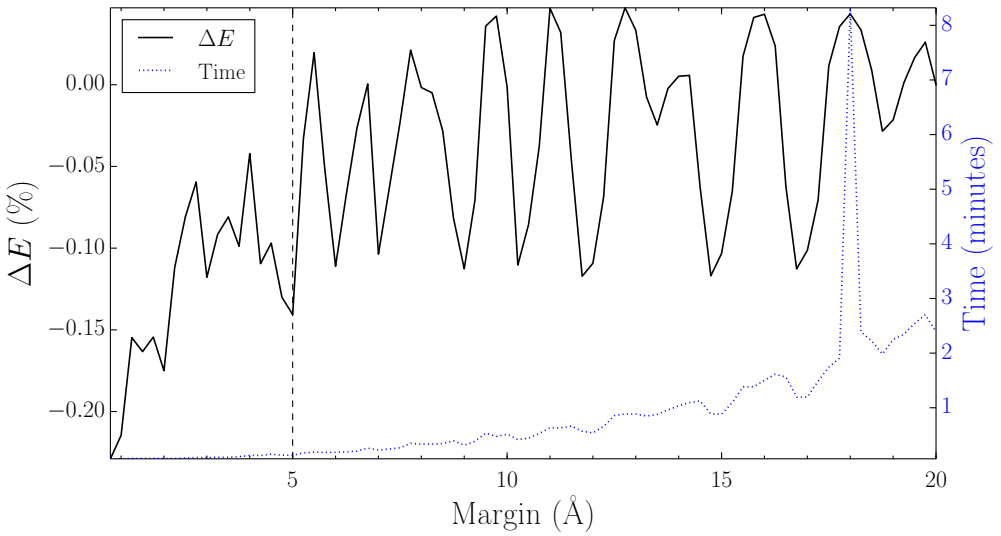


Figure A.3: The energy difference and elapsed time for different margin values compared to the last tested margin value of 20 Å. The vertical line indicates the value 5 Å chosen for the simulations. The test system consists of one sodium surrounded by six closest water molecules extracted from a classical MD snapshot.

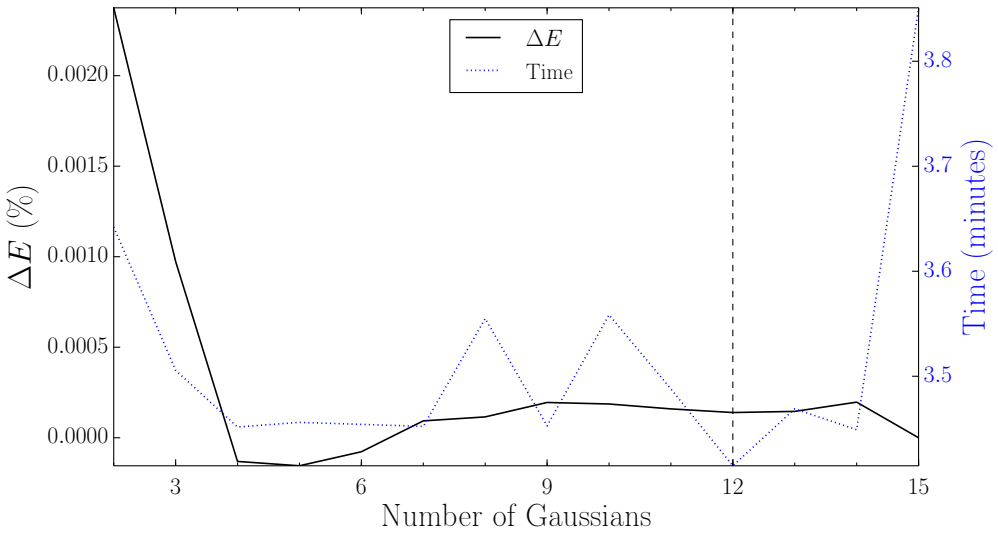


Figure A.4: The energy difference and elapsed time for different number of Gaussians used in the GEEP expansion compared to the highest available number of 15 gaussians. The vertical line indicates the value 12 chosen for the simulations. The test system corresponds to the one used in the final QM/MM simulations, namely 1000 water molecules and one sodium where positions are extracted from classical MD.